Univerzita Karlova v Praze
Přírodovědecká fakulta

Studijní program: Biologie
Obor: Teoretická a evoluční biologie

Bc. Adam Klimeš

Community ecology from the perspective of classic and Bayesian statistics
Ekologie společenstev z hlediska klasické a Bayesovské statistiky

Diplomová práce

Vedoucí práce: Mgr. Petr Keil, Ph.D.

Praha 2016

Poděkování:

Na tomto místě bych rád poděkoval za odborné konzultace, podporu a připomínky k předchozím verzím práce Jitce Klimešové a Petru Keilovi.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 15. 8. 2016

.....

Adam Klimeš

Abstract

Quantitative evaluation of evidence through statistics is a central part of present-day science. Bayesian approach represents an emerging but rapidly developing enrichment of statistical analysis. The approach differs in its foundations from the classic methods. These differences, such as the different interpretation of probability, are often seen as obstacles for acceptance of Bayesian approach. In this thesis I outline ways to deal with the assumptions of Bayesian approach, and I address the main objections against it. I present Bayesian approach as a new way to handle data to answer scientific questions. I do this from a standpoint of community ecology: I illustrate the novelty that Bayesian approach brings to data analysis of typical community ecology data, specifically, the analysis of multivariate datasets. I focus on principal component analysis, one of the typical and frequently used analytical techniques. I execute Bayesian analyses that are analogical to the classic principal components analysis, I report the advantages of the Bayesian version, such as the possibility of working with uncertainties and missing values, and I highlight its ability to make predictions. Although not yet fully developed, Bayesian approach promises new possibilities to community ecology; these will likely alter the community ecology as we know today. Apart from the new methods handling missing values, uncertainties, and predictions, there is also the opportunity for improvement of the classic approach, based on the lessons from Bayesian approach.

Abstrakt

Významnou součástí dnešní vědy je kvantitativní zhodnocení dat prováděné statistickou analýzou. Bayesovský přístup představuje začínající, ale rychle se rozvíjející, obohacení statistické analýzy. Liší se ve svých východiscích od klasických metod a tyto rozdíly, jako například odlišná interpretace pravděpodobnosti, jsou považovány za překážky k přijetí Bayesovského přístupu. V této práci vyznačuji cestu, jak přistoupit k předpokladům Bayesovkého přístupu a jak se postavit k hlavním námitkám vůči němu. Bayesovský přístup prezentuji jako nový způsob analýzy dat odpovídající na otázky, které si ve vědě klademe. Bayesovský přístup hodnotím z pozice ekologie společenstev. Na příkladech ilustruji nové možnosti, které tento přístup nabízí pro analýzu mnohorozměrných dat typických pro ekologii společenstev. Zaměřuji se zejména na často používanou analýzou hlavních komponent. Představuji její Bayesovskou analogii a výhody, které přináší oproti klasickému přístupu, jako je například práce s nejistotou odhadů. Bayesovská analýza mnohorozměrných dat v ekologii je teprve ve svých počátcích, nicméně zavádí cenné nové možnosti pro ekologii společenstev. Využití Bayesovského přístupu povede ke změně ekologie společenstev, jak ji dnes známe. Ekologii společenstev se nabízí nejen nové metody umožňující efektivně pracovat s nejistotami, s chybějícími hodnotami a vytvářet predikující modely, ale navíc osvojení Bayesovského přístupu představuje příležitost pro reflexi klasického pojetí a zdokonalení používaných metod.

Keywords

Interpretation of probability, philosophy of science, Bayesian statistics, community ecology, multivariate analysis, principal component analysis

Klíčová slova

Interpretace pravděpodobnosti, filosofie vědy, Bayesovský přístup, Bayesovská statistika, ekologie společenstev, mnohorozměrná analýza, analýza hlavních komponent

Contents

1. Introduction

Since the late nineties and thanks to advances in computational tools, there has been a rising interest in Bayesian statistics, with a notion that environmental scientist are becoming Bayesians (Clark, 2005). In line with this development, the title of this thesis could have been: "What does it mean to become a Bayesian for a community ecologist?" It is not the case, since one does not necessarily adopt a completely new paradigm or new methodology, which would require abandonment of the current beliefs and methods. Although the use of Bayesian approaches has been on the rise, here I aim to show that they can work alongside the classic methods (Clark, 2007). Hence, I hope to present Bayesian approach as a set of new possibilities in data analysis, which could improve our ability to deal with various challenges.

My target scientific field will be community ecology – the study of patterns in the diversity, abundance, and composition of species in communities, and of processes underlying these patterns (Vellend, 2010). Community ecology often deals with multidimensional, incomplete and error-infested datasets, which need sophisticated analytical methods for both the evaluation of particular hypotheses as well as for exploratory analysis. I will show that Bayesian statistics introduces techniques which can manage problematic datasets not analyzable with classic methods, and it can provide results which could not be acquired otherwise. Although Bayesian statistics developed its foundations more than a century ago, it is still not used as much as it could. There are two groups of reasons: practical – as limited implementation of Bayesian analyses in popular software (Breslow, 1990), and theoretical – as different interpretation of probability and assumptions which go with it, making together the broader adoption of Bayesian methodology nontrivial (Ellison, 2004). In this thesis I hope to provide a manual for such step, and I will attempt to address questions: "What does it mean to use Bayesian approach in community ecology?" and "Will adoption of Bayesian approach change community ecology?"

1.1 Community ecology

Community ecology is a scientific field half way between population ecology and macroecology. It focuses on communities, assemblages of organisms on particular habitat in particular time (Vellend, 2010). Community ecology studies interactions between these organisms, their origins, and their ecological and evolutionary consequences (Cavender-Bares, Kozak, Fine, & Kembel, 2009). The diverse object of the study raises a question whether there could be any general patterns or laws, or whether community ecology is predestinated to be

highly reductionist and local in focus (Lawton & Kinne, 2000; Simberloff, 2004). Recently, there has been a decline of the reductionist and traditional approach in community ecology, where dynamics of small number of species has been studied (Lotka, 1910). The field has been shifting towards analyses of functional traits (McGill, Enquist, Weiher, & Westoby, 2006) and towards large spatial scales focusing on statistical emergent properties of the entire system, rather than its individual components. The object of interest is not a couple of species any more, but rather a whole community. This approach inevitably brings multidimensional datasets; however, such data in community ecology are often incomplete, infested with measurement errors, and biased. These are issues that are well handled by Bayesian statistics.

1.2 Statistics as methodological tool

In science, knowledge about general patterns is presented in the form of theories, usually represented by sets of hypotheses or models, which we evaluate in the light of evidence, commonly called data. The evidence is acquired through observation or experimentation; the evaluation is done using methods that should be as objective as possible (Dawid, 2004), and such method in modern science is statistics. It enables to infer general patterns or laws from the limited evidence, and helps to distinguish between systematic (deterministic) effects of interest and random variability which is usually not of interest per se. There are several paradigms in statistics, each of them encompassing different assumptions and interpretations (Breslow, 1990). One such paradigm is Bayesian statistics.

1.3 Bayesian statistics

Bayesian statistics is an inferential tool which uses *Bayesian confirmation theory* (Crupi, 2015) to evaluate the impact of available evidence on the hypotheses. Bayes´ theorem, the central pillar of the confirmation theory, is named after Reverend Thomas Bayes, and it was introduced in famous paper: “An Essay towards Solving a Problem in the Doctrine of Changes” (Bayes & Price, 1763). The theorem states that conditional probability of a hypothesis given the evidence is proportional to probability of the evidence given the hypothesis, multiplied by prior probability of the hypotheses (see more in section 3). This serves as a basis for a confirmation theory which differs from the classic one.

Bayesian statistics is still not widespread, but is on the rise in ecology; however, it is nearly missing in community ecology and in other fields that are characteristic by heavy use of

multivariate analyses. Because Bayesian statistics has a lot to offer for multivariate analysis, I will present how it can be applied to multivariate data and to the corresponding questions.

1.4 Plan of the thesis

Having mentioned the classic and Bayesian approaches, we can zoom out to philosophy of statistics, or specifically to assumptions summarized in confirmation theory and to the very interpretation of probability. All of these topics fall into the general field of philosophy of science (Romeijn, 2016), which asks what knowledge is and how it is acquired. I will touch these broad topics in this thesis, but I will not address them systematically.

I will begin with clarification and comparison of the terms *classic* and *Bayesian* – this is necessary, since the terms are somewhat ambiguous. At the same time, I will present interpretations of probability which can be denoted as classic or Bayesian, and which are tightly connected with classic and Bayesian statistics respectively. To complete the picture, I will describe the most frequently discussed differences between the two kinds of statistics. I will then discuss how the philosophical differences between the classic and Bayesian statistics affect practical data analysis.

In the final part I will move entirely to the practical issues: I will focus mainly on analyses of multivariate data which are common in community ecology. I will compare classic and Bayesian statistics using an example of real-world[1] data analysis. Since multivariate Bayesian data analysis is still in its infancy, I had to develop several tools, especially visualization R functions which allow to fully appreciate the advantages of Bayesian approach. Finally, I will summarize the main advantages and improvements which Bayesian approach offers to the classic data analysis. I will provide R code for the practical part in the appendix.

[1] These are data about plants. Correspondingly I will focus on community ecology dealing with plants. On the other hand, majority of this thesis is independent of primary object of study.

2. Classic and Bayesian interpretation of probability

Here I will delineate the meaning of the terms "classic" and "Bayesian", since they refer to a wide range of ideas. The two terms can label concepts on different levels of generality; at the lowest level, they can label the very interpretation of probability.

2.1 Classic interpretation of probability

Classic interpretation of probability can denote two concepts, which have a lot in common, but are distinct. Those are classic interpretation, as we will call it, and frequentist interpretation of probability. The classic interpretation of probability refers to probability based on games of chance – an idea that was introduced in works of Laplace, Pascal, Bernoulli, Leibnitz, and others (Hájek, 2012). In the classic interpretation, all events are reduced to equally possible cases[2], so that we may be equally undecided in regards to their existence. Probability is then the ratio of favorable cases to all possible cases (Laplace, 1902). There are several critical points which have to be further elucidated in this definition. First, referring to "equally possible cases" seems to presuppose the notion of probability, and thus the definition is circular[3]. Second, it follows from the definition that probabilities can only have certain values – probabilities are defined as ratios, so only rational probabilities are possible. Third, extension of the definition to infinite space can be done, but we lack unequivocal way of determining our indifference in regard to the existence of possible outcomes. This is illustrated by Bertrand´s paradox (see section 3.1): there is not always a straightforward way to determine our indifference to the space of possibilities.

2.2 Frequency interpretation of probability

The frequency interpretation is the one that is most widely used, and therefore it is often designated as the classic probability. It has two interpretations: Finite and hypothetical frequentism. *Finite frequentism* (Venn, 1876) is similar to the classic interpretation: probability of an event is the relative frequency of actual occurrences of the event in a reference class[4]. It means that we (again) assign equality to a set of events, and we derive the probability as a proportion or frequency of favorable cases. But in the case of finite frequentism, we speak about actual cases, not about all possible (hypothetical) cases. Apart from the problems which this

[2] Also referred to as *outcomes* in statistical literature.
[3] See Hájek (2012) for further discussion of whether this circularity can be avoided.
[4] Probability of particular meadow being invaded by *Heracleum mantegazzianum* is relative frequency of meadows invaded by this species. The reference class are meadows.

interpretation shares with the classic one, there is the so called "single case problem": If we have only one outcome, then its probability according to finite frequentism is its relative frequency, which is one (Hájek, 2012), and that is not how we usually understand probability. The problem indeed proliferates to two cases, three cases, and so on. With two cases the possible value of probability is 0, ½, and 1; with three cases we have 0, 1/3, 2/3, and 1.

Hypothetical frequentism (Reichenbach, 1971) attempts to address the problems mentioned above: instead of the actual frequency of cases, probability is defined as a limiting frequency. Limiting frequency is frequency of infinite number of trials. However, there is nothing corresponding to the limiting frequency in the physical world, because we have usually only limited number of cases/trials. We have to imagine infinite extension of an actual sequence of trials; probabilities are then frequencies of outcomes of this hypothetical infinite sequence of trials. The term is not clear, since it can be objected that certain events are in principle unique[5], and so there is nothing to build the limiting frequency on. Even in the case of non-unique events there is a question of what is the reference class to which we should determine the limiting frequency[6].

2.3 Subjective interpretation of probability

The term "Bayesian" is associated with *subjective interpretation of probability*, which defines probability as a degree of belief. Often stated as willingness to accept certain bets: "Your degree of belief in E is p, if p units of utility is the price at which you would buy or sell a bet that pays 1 unit of utility if E, 0 if not E" (Hájek, 2012). As such degree of belief does not refer to anything physical, just to a confidence of suitable subjects called agents. There are different constrains put upon these agents and their beliefs. A radical subjectivist position would let the agents believe whatever they want. Mainstream interpretation would use the *Dutch book argument* to constrain beliefs of agents in the way that they follow axioms of probability[7] (Kolmogorov, 1956). A Dutch book against an agent is a series of bets which are acceptable (according to the quoted definition) to the agent, but which lead to a certain loss on his side. Agents that disrespect the axioms of probability are susceptible to the Dutch book, agents who follow the axioms are not (Kemeny, 1955). This means that it is possible to design a series of bets leading to their

[5] e.g.: Result of EURO 2016 football championship, assassination of Reinhard Heydrich

[6] Let's return to our example with *Heracleum*: The meadow of interest is small, on high slope, sometimes fertilized. Now comes the question what is the probability of it being invaded. Is it probability of meadow as small? That would be frequency of invaded small meadows. Or is it probability of meadow on high slope or as fertilized? There are different probabilities according to which reference class we choose.

[7] Interestingly, there are no universally accepted axioms of probability. The most frequently used where introduced by Kolmogorov in his *Grundbegriffe der Wahrscheinlichkeitsrechnung* in 1933 (Kolmogorov, 1956).

certain loss. Beliefs which are susceptible to the Dutch book can be marked as incoherent and irrational. There are further constrains which can be put on agents and their degrees of belief. Namely, their beliefs should be updated only according to Bayes' theorem.

We can raise several objections to the subjectivist/Bayesian interpretation of probability, and such objections could prevent the acceptance and practical applications of Bayesian approach – therefore the objections are critical for the purpose of this thesis. I will discuss them in more detail in the following section.

2.4 Critique of the subjective interpretation of probability

The first objection against the subjective interpretation of probability is aimed at the very word "subjective": In science we seek objective knowledge[8], so how can we justify the use of the subjectivist probability? It can be argued that if we constrain the subjectivist probability by probability axioms and updating rule, then the probability will behave independently of the subject. This is often called *objective Bayesianism*.

If we go back to the subjectivist interpretation of probability (section 2.3), we see that probability has been defined as a willingness to bet on certain outcomes. There are problems with this: In Bayesian statistics we assign probability to hypotheses, but some hypotheses are practically impossible to be verified or falsified. Does it make sense to bet on such hypothesis? To this group we can also include all probabilistic hypotheses, which we usually deal with in science. Further, it can be questioned whether it makes sense to identify beliefs with willingness to bet on them – we can imagine that we believe in something, but we have no interest in betting on it. Typical example is willingness to bet on our own death (Romeijn, 2016).

It can be argued that these problems arise because of taking the definition literally. Constraining believes by axioms of probability and updating rule suggests that we are dealing with some idealized degrees of belief. Those are beliefs of some imaginary rational agent rather than of an actual human being. Probably some reformulation in the form of Rawls' "Veil of ignorance"[9] would be more instructive (Hájek, 2012; Rawls, 1971).

[8] Emphasizing of objectivity in knowledge can be seen as pleonasm. Knowledge as justified true believe (Plato, 1883) is necessary objective.

[9] John Rawls in his famous book "A Theory of Justice" introduces a thought experiment "Veil of ignorance" while developing the concept of justice. Shortly: An organization of society can be considered as "just" if it is approved by rational agents who have no knowledge about their position in the society.

2.5 Meaning of "classic" and "Bayesian"

Finally, I shall note that there is yet another understanding of the term "classic". It does not concern the interpretation of probability, but rather the particular methods of data analysis. As "classic" we can denote the methods that are used often and are typical in a given scientific field. The classic methods would then contrast with "novel" methods, a category that would also encompass Bayesian statistics.

In this thesis my goal is to compare Bayesian approach with the classic one, with emphasis on what exactly does it mean to start using Bayesian statistics in data analysis, particularly in community ecology. Following the example of Jaynes (1976), Clark (2007) and Bolker (2008) I employ a pragmatic attitude, and I focus on the parts which have impact on data analysis. There are the multiple interpretations of probability with even more subtle details, yet some of them are irrelevant for practical analysis. Hence, my stance is same as presented by Dawid (2004): metaphysical positions which are irrelevant to our conclusions are not of interest. I argue that it does not matter for our purpose whether we believe in a strict determinism and understand probabilities only as a measure of lack of information about particular system, or we see them as proprieties of systems – so called propensities[10] (Popper, 1959) as long as there would be no practical difference between those two positions.

With this in mind I will use the term "classic" in the sense of the typically used methods. In the case of univariate data analysis, classic statistics will be used to denote methods assigning probabilities to some sample space (population from which our samples originate). Such probabilities can be interpreted as frequencies. Classic methods are then used to eliminate hypotheses which mark our sample case(s) as improbable (Romeijn, 2016). This position is usually associated with frequentist interpretation of probability, although it is compatible with different interpretations – classic, propensity, or even subjectivist. These metaphysical viewpoints would alter our presentation of the results, but would not prevent our usage of classic statistical methods. Importantly, the presentations of results would differ in vocabulary among the viewpoints, though numerically and practically there would be no difference.

In the case of multivariate data analysis, there is a wide use of various ordination techniques. Some of them can be seen as mere data transformations (such as principal component analysis), and as such are completely independent of probability interpretations.

[10] According to propensity theory of probability there is something in physical world corresponding to our theoretical terms like P(x). In particular situation, for example, coin has certain propensity to fall head up. This propensity is completely independent of our theorizing (Dawid, 2004; Popper, 1983).

When I speak about classic methods in multivariate analysis, I refer to these data transformations.

I will use the term “Bayesian” to refer to Bayesian inference, with the associated subjectivist interpretations of probability, and with the corresponding statistical methods.

3. Bayesian inference

Bayesian statistics is built on Bayesian inference, and Bayesian inference works with subjectivist interpretation of probability. As outlined in section 2.3, the subjectivist interpretation views probability as a degree of belief. We can assign such probability to hypotheses[11]. Bayesian inference states how we update our degree of belief in a particular hypothesis in the light of new evidence. This is done using Bayes' theorem:

$$P(h|e) = \frac{P(e|h) * P(h)}{P(e)}$$

$P(h|e)$ marks conditional probability – probability of h (the hypothesis) given e (the evidence). $P(e|h)$ is usually referred to as *likelihood*. $P(h)$ is *prior* probability of hypothesis, and $P(h|e)$ is the posterior probability of the hypothesis. From Bayes' theorem we can derive several rules that are important for inference (for more details see Joyce (2008)):

1) In case $P(e) = 1$, i.e. when the evidence was predicted as certain by all possible hypotheses, we get $P(e|h) = 1$. From this follows that $P(h|e) = P(h)$. Such evidence has no influence on probability of the hypotheses.
2) From Bayes' theorem we can say when the evidence confirms[12] the hypothesis, and we can determinate a measure of the confirmation. Specifically, the evidence confirms the hypothesis when $P(e|h) > P(e)$ (Rosenkrantz, 1983). The difference or proportion of $P(h|e)$ and $P(h)$ were suggested as the measure of confirmation (Crupi, 2015).

3.1 Problems with Bayesian inference

Bayesian inference was shortly presented in the previous section as updating of our degree of belief in a hypothesis in the light of a new evidence. The controversial part is the use of prior information in this inference.

Prior information means probability (the degree of belief) in a hypothesis before we encounter the evidence[13]. Here we again confront subjectivity – prior information is subjective, so the posterior probability must also be to some degree subjective, which is in conflict with the objectivity we require from scientific inference.

[11] In the classic statistics there is a probability of getting particular data under the assumption that our hypothesis holds, but the hypothesis itself can only be true or false. It does not make sense to probability to such hypothesis under this interpretation.
[12] Confirmation does not of course mean decision whether hypothesis is true or false. Evidence just updates the probability of hypothesis.
[13] Term "prior" denotes our degree of belief, before we encounter an evidence, not before all evidence. Prior does not mean a priori; in fact, as it will be mentioned, posterior probability from a previous analysis can be used as a prior in a following analysis.

Let me consider two cases of specification of prior information: We either may know something about a hypothesis before we get the evidence, or we may not. First, in case we have some information, the prior probability should[14] encompass it. If the information comes from a previous study as a posterior probability, then we simply use it as a prior in the new analysis. The second case is tricky: we need to quantify our ignorance. This can be done, for example, by using a prior which assigns the same probability to all possible hypotheses. For continuous parameters this means that they have probability density uniformly distributed over all possible parameter values[15]; this is sometimes called a *flat* or *uninformative* prior (Gelman, Carlin, Stern, & Rubin, 2004). Yet specification of uninformative priors can be problematic, since it faces the so called *Bertrand's paradox* (Bertrand, 1889).

Bertrand's paradox illustrates ambiguity in assigning probability[16]. Consider we are asked about probability that a randomly located chord within a circle is longer than a side of an equilateral triangle inscribed in the circle. Bertrand offers three ways of determining this probability (Fig. 1):

1) We choose two random points on the circle. We then rotate the triangle in a way that one of its vertices coincides with one endpoint of the chord (Fig. 1A). The chord is longer than the side of the triangle if its other endpoint lies between the other two vertices of the triangle. Length of this arc is one third of the circumference of the circle, so the probability is 1/3.
2) If the radius of the circle is made perpendicular to one side of the triangle, and then we construct chords that are perpendicular to this radius, we can see that the chords intersecting the radius within the triangle are longer than its side, and the chords outside of the triangle are shorter (Fig. 1B). Intersection of the triangle and the radius lies at the half of the radius, so the probability is ½.
3) We construct chords by choosing the points within the circle as their midpoints first. Chords with their midpoint within the circle which is inscribed into the triangle (green circle in Fig. 1C) are always longer than side of the triangle. Otherwise they are shorter. Area of the large circle is four times larger than area of the small one. The probability is ¼.

[14] Lower is suggested, that there are cases, when we would rather not use prior information.
[15] In case of infinite range of particular parameter, we do not use uniform probability distribution, because it assigns zero probability to all values. Instead we approximate it with wide normal distribution.
[16] Jaynes offers solution of this paradox, but it is not universally accepted (Jaynes, 1973).

Surprisingly, all three presented ways are equally right for the determination of the probability! The problem – what is the probability – does not have a unique solution, unless we somehow specify the process of random positioning of the chord within the circle.

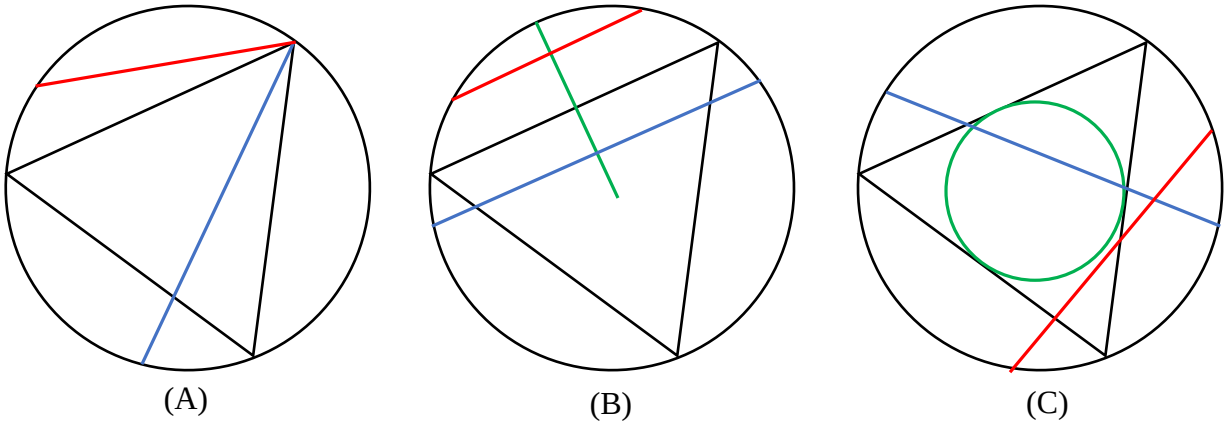

Fig. 1: Illustration of Bertrand's paradox. (A) First case assigning probability to the chord being longer than the side of the triangle 1/3. (B) Second case assigns probability 1/2 and (C) the third case assigns probability 1/4. Blue chords are examples of those longer than the side of the triangle. Red chords are examples of those shorter than the side of the triangle. Green color is used for auxiliary structures.

One solution of the problem of subjectivity of the priors comes from praxis: The impact of prior information weakens quickly with increasing amount of evidence, resulting in convergence of posterior distributions that were obtained by using different priors (Howson & Urbach, 1991). Another response points out that the subjective decisions are always present in any statistical analysis (Van Dongen, 2006): As an example of a subjective element in classic statistics we can mention the so called *stopping rule problem* (see section 4.1).

Finally, there is critique of the whole process of updating of probability – known as *the problem of old evidence*, there is a question of how to evaluate impact of evidence older than the theory. Can such evidence be used to confirm the hypotheses or not? One possible solution is to simply postulate that the timing of evidence and theory is arbitrary (Crupi, 2015). An alternative solution is to avoid the problem by restricting the evidence we use in the analysis.

Bayesian approach allows us to combine our information about the hypotheses with new evidence. In the classic approach we do not use previous information in analysis. Analogically we can neglect available information or part of it in Bayesian approach. Of course then the posterior probability will not express degree of belief in a particular hypothesis given all available information, but just given the information that we used. This way we can avoid problems with including evidence that we are not able to quantify as prior distribution or we do not want to (for example because of the problem of old evidence).

4. Classic and Bayesian statistics and data analysis

There are several points on the level of statistics and data analysis which deserve extra attention. We saw that classic and Bayesian approach differ in the interpretation of probability, and they also do confirmation differently. Some of these differences proliferate into practical data analysis. I focus on those which are frequently encountered.

4.1 Stopping rule problem

In the classic approach the result of an analysis is usually (in the case of hypothesis evaluation) probability of the data given the hypothesis. However, this is imprecise. This probability, the so called p-value, is probability of our data, *or more extreme data*, given the hypothesis (McCarthy, 2007). For example, we have 100 pots with soil and we plant a single *Poa annua* seed in each of them. After few days we find that only 40 seeds germinated. Now what is the probability that a colleague's hypothesis, based on years of experience, that the germination rate is 50%, is right? Binomial test gives the p-value of 0.057. This is not the probability of getting exactly 40 germinated seeds; it is the probability of observing such and more extreme data given the hypothesis that the germination rate is 0.5. So it encompasses data ranging from 0 to 40 and from 60 to 100 germinated seeds (Figure 2).

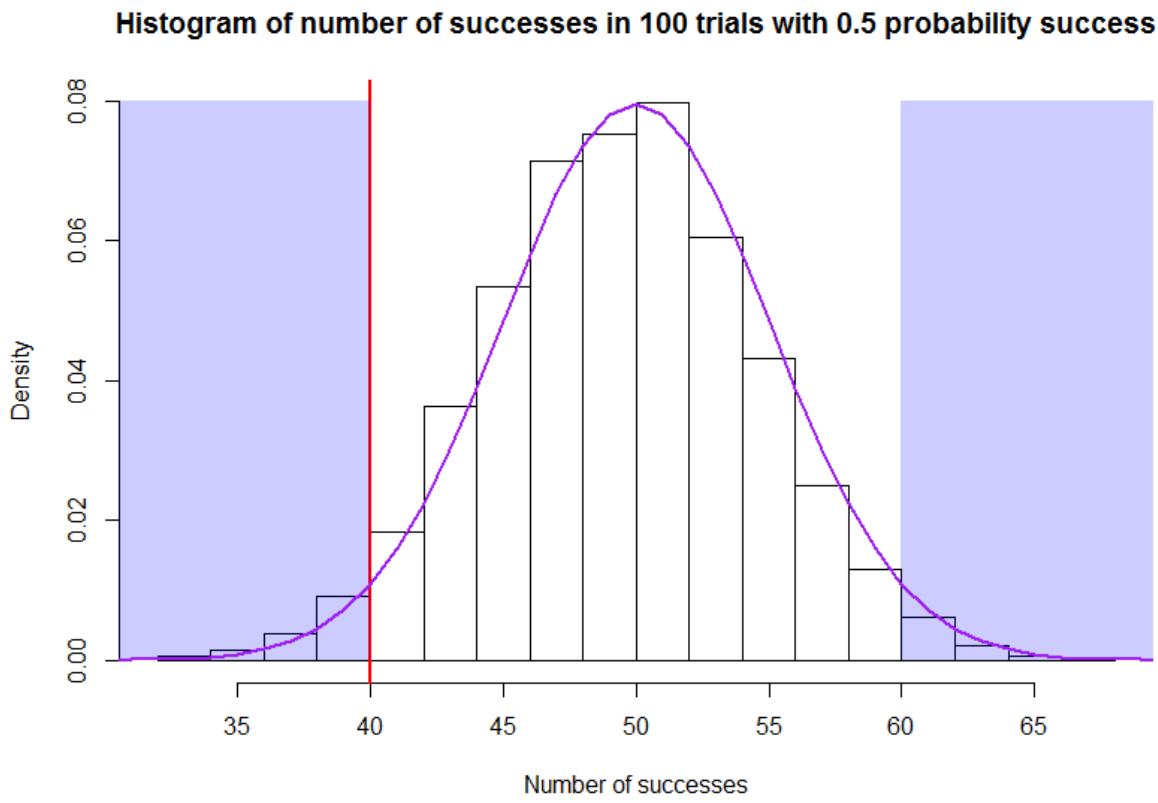

Fig. 2: Histogram illustrating the cases which are evaluated by p-value. The histogram is constructed from 10000 repetitions of series of 100 trials with probability of success 0.5. The result from the example (40 successes) is marked by red line. Purple line is a theoretical binomial probability density function for probability 0.5 and 100 trials. P-value of our result denotes probability of getting any number of successes which lies in the blue area.

The stopping rule problem arises when we alter the procedure. This time we plant more than 100 seeds (each in a separate pot) and after some time we check the germination success in each pot. However, now we count the number of germinated seeds until we reach 60 empty pots. We end up with 40 germinated seeds – exactly the same result as before – but our p-value is now 0.058[17].

The reason why the p-values differ is that there are different data we could get. As I said, the p-value expresses not only the probability of the actual data, but also data more extreme under a particular hypothesis, and those differ between the two procedures. In the first case we could never find more than 100 germinated seeds, but that is possible in the second case. Figure 3 shows the approximated distribution for the second procedure; it is slightly positively skewed.

[17] In the example the p-value is similar in both procedures, suggesting that the stopping rule has a negligible effect. On the other hand, due to the common emphasis on whether the p-value is under or above 0.05, the example could be easily constructed in such way that the stopping rule makes a difference between significant and insignificant result. For such examples see: (Howson & Urbach, 1991; Lindley & Phillips, 1976; McCarthy, 2007).

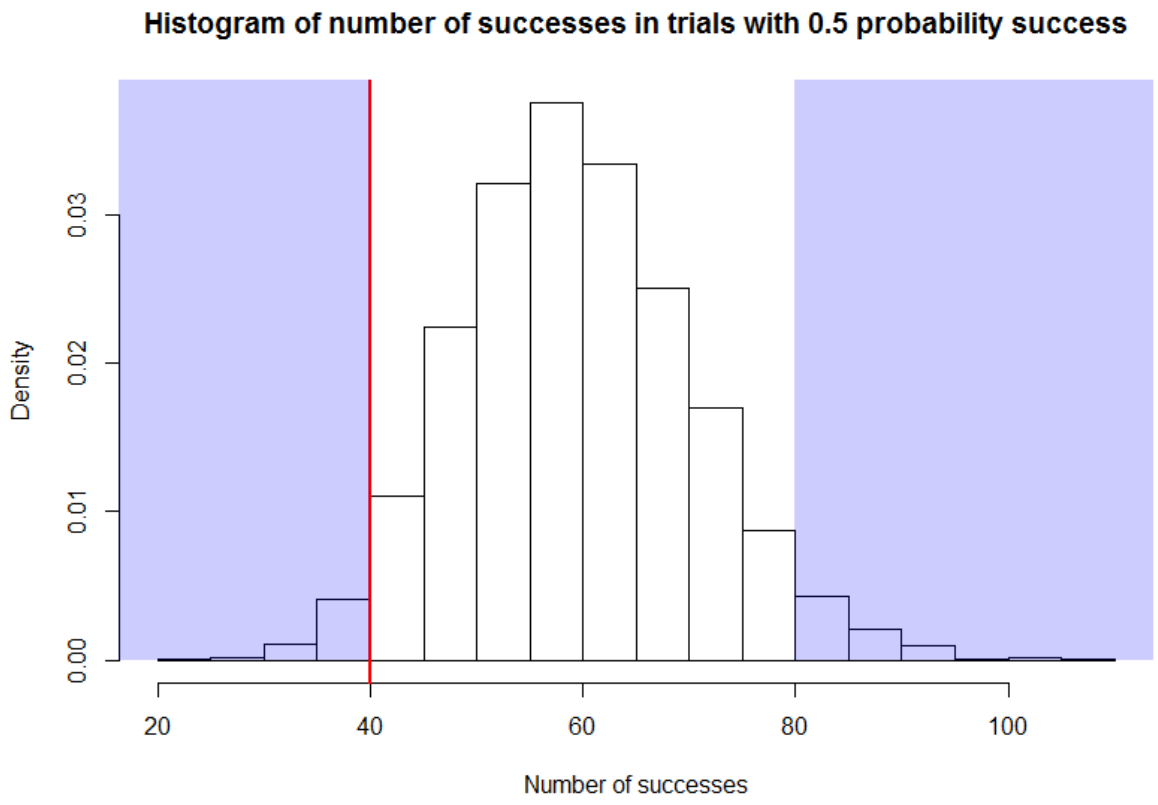

Fig. 3: Histogram approximating the distribution of results in case of the second stopping procedure. The histogram is constructed from 10000 repetitions of a series of trials. Each series consists of trials with 0.5 probability of success and stops when it reaches 60 fails. Red line marks the result (40 successes) from the example. Blue area marks the number of successes to which the p-value refers.

Dependency of results on the stopping rule is often viewed as a weakness of the classic approach (Howson & Urbach, 1991). It introduces, or can introduce, subjectivity in the data analysis, since the procedure of the data collection can easily be forgotten, missed out in the data analysis, or not even reflected in the first place (Lindley & Phillips, 1976). On the other hand, it can be accounted for, and with careful experiment preparation we can minimalize the risk of wrong inference.

The stopping rule problem is used as an argument that scientific inference inevitably encompasses subjective decisions, and as a case for the use of prior probabilities in Bayesian statistics (Ellison, 2004). Some even argue that because any inference unavoidably contains subjective elements, it impossible to eliminate them all[18].

[18] It could be argued even further that subjective decisions in science are beneficial. It is possible to show, on historical examples, that scientific inference often does not obey methodologies introduced by philosophers of science, and such behavior can actually benefit science (Feyerabend, 1993).

4.2 Confidence vs. Bayesian intervals

When we estimate a particular parameter value in the classic approach, we also report confidence intervals to express uncertainty in the estimate. Usually 95% confidence intervals are given, corresponding to conventional 5% error of refusing the true hypothesis. Confidence intervals (95%) mean that, with repeated sampling, 95% of the time our confidence intervals would encompass the true (population) value which we are estimating (Neyman, 1937). Hence, this particular confidence interval does not refer to the probability of the true value being in the interval, neither that it covers the true value, as it is often interpreted (Havránek, 1993).

In Bayesian inference we report the uncertainty of the estimation in the form of Bayesian (credible) intervals. Here we can speak about the probability of the true parameter value being in the interval. This is possible, since under Bayesian approach we assign probabilities to parameters. The posterior probability density (or mass) is exactly this – it gives the uncertainty about a particular parameter, estimated on the basis of the data and the prior information.

4.3 Small datasets

Data in community ecology are usually laborious and expensive to obtain. From this follows the necessity for optimization of experiments in order to fit to our budget, and we need to have sufficient sample size in order to reach a desired test power in the data analysis. Either way, we will probably end up wishing for more samples, so that we can prove or disprove the weaker effects in our data.

Second common case of having too small datasets (to test hypotheses) is in case of pilot studies. In community ecology we often need to try new methods on small samples first, to see whether everything works before we start a larger experiment. If we are lucky, results from such pilot study are usable.

The classic approach does not allow to make use of small datasets. To test hypotheses, our dataset must be large enough, or it has no value. Extremely small datasets (e.g. N = 2) are not evaluable at all, since we do not have enough observations to estimate parameters of our model[19]. Bayesian approach, on the other hand, can evaluate even very small datasets. In such case the posterior distribution will be similar to the prior one as we would intuitively expect. However, this posterior distribution still has a value, since it can be used as prior in following studies.

[19] For example, if we have a single observation and want say something about mean of a normally distributed population, we cannot. In terms of degrees of freedom, we have one, which can be thought as used on (wasted on) estimation of the mean, and there is no residual variability and no residual degrees of freedom.

5. Impact of the theoretical background on practical data analysis

5.1 Interpretations of probability

So far I have introduced several interpretations of probability, and the main objections that have been raised against them. These were the classic interpretation, frequentism, hypothetical frequentism, and subjective interpretation of probability[20] (and the briefly mentioned propensity interpretation). I will now try to answer the question: "Do any of these interpretations limit our practical usage of tools of statistical analysis?"

In reality the problem often goes the other way around: the criterion we would impose on interpretation of probability is to explain probability in the way we use it. We need the specific interpretation of probability to be the basis of the statistical methods that we use. Any interpretation of probability which would fail to consider our basic statistical methods would be discarded[21]. Jaynes (1976) suggests that the merits of any statistical method are determined by the results it gives when applied to specific problems, or to also quote Jaynes literally: "*The merits of any statistical method are not determined by the ideology which led to it.*" From this follows that the necessary criterion for acceptance of an interpretation of probability is its compatibility with the statistical methods which give results that are accepted by our commonsense judgment[22].

I assume that an ecologist's interpretation of probability is compatible with the methods that are universally used among ecologists. So whatever the basis of a particular interpretation we are dealing with, be it classic, frequentist, subjective etc., I only need (for purpose of this thesis) to ask whether it is also compatible with the Bayesian approach.

Bayesian approach has one demand on the interpretation of probability which is not shared with the classic approach. While in Bayesian analysis we assign probability to hypotheses and their parameters, in the classic analysis we take hypotheses as true or false, and parameters as given, although often with unknown values. From this follows that our interpretation of probability, in order to be compatible with Bayesian analysis, has to allow assigning probability to hypotheses (and parameters). Without this possibility Bayesian analysis does not make sense. Bayes' theorem, which governs Bayesian analysis, makes sense only if

[20] Mentioned interpretations of probability have plenty of variants and so can be considered more like groups of similar interpretations. Differences between variants go beyond the scope of the thesis. For literature to this topic see: (Hájek, 2012).

[21] Some authors consider their probability interpretations as those used in practice (Popper, 1959).

[22] Commonsense judgment is here used in wide meaning. As Jaynes pointed out we are not often able to distinguish between different statistical methods on basis of our common sense. Numerical results could be nearly identical. We need to find an example, where these small differences in results would be greater or where they would lead to distinct qualitative conclusions. Then we can tell which method is preferable.

we allow the existence of probability of hypotheses[23]. Among the interpretations of probability mentioned earlier, the only one which explains meaning of "probability of hypotheses" is the subjective one. Probability of hypotheses is the degree of belief in them.

5.2 Statistical inference

I have already mentioned the danger of subjectivity which can creep into both approaches: Bayesian approach uses prior information and classic approach faces the stopping rule problem.

Prior information can be used in a way that does not compromise the objectivity of Bayesian analysis: we can take prior results from previous studies, or we use uninformative priors. In the former case we can encounter *the problem of old evidence*. The minimalistic solution of this is to limit evidence that we evaluate, as was suggested in section 3.1. While using uninformative priors, on the other hand, we face *Bertrand's paradox* – the ambiguity of exact specification of our ignorance. However, *Bertrand's paradox* is, from the perspective of practical data analysis in community ecology, a theoretical problem. The specific distributions that are typically used, e.g. wide normal distributions in case of continuous parameter values, usually do not have distinct "flat" alternatives that would also led to distinctly different posteriors[24].

Further, the stopping rule problem is solvable. We just have to pay attention to it when we collect our data, and it needs to be properly acknowledged in our analysis. I argue that this should at least weaken the stopping rule problem, if not entirely eliminate it.

There are similar problems in the classic analysis. As Jaynes (1976) illustrates on examples, various analyses can be elegantly done using Bayesian approach and not the classic one. On the other hand, problems reported with the classic approach are often caused by the absence of established methods for a particular type of analysis, or caused by errors which could have been avoided.

5.3 Statistical analyses

In section 4.3 I have mentioned differences between the approaches in their ability to deal with small datasets. From what was said it directly follows that the two approaches differ in their usability for particular datasets. The advantage of Bayesian approach is the ability to deal with

[23] Of course Bayes' theorem, as rule of conditional probability, still holds without this condition, but it cannot be interpreted in the context of evidence and hypotheses.

[24] This does not mean that we should not pay attention to the determination of priors. As Van Dongen (2006) shows, priors can have great impact (in case of small samples) on our results and their determination is not trivial.

small datasets. On the other hand, large datasets are often problematic, because their analysis can be computationally intensive and difficult. Indeed, what we perceive as "large" data (also colloquially termed "big data") changes quickly with increasing power of computers; nonetheless, the need for efficient algorithms remains, since some analyses are impossible to execute in practice ever[25].

The other end of analysis, the reporting of the results, also differs between the classic and Bayesian approach. As I have briefly outlined in section 4.2, the interpretation of confidence and Bayesian intervals of the estimated values is not the same. However, one can argue that, when uninformative priors are used, the difference between Bayesian intervals and confidence intervals is not practically important (Clark, 2005). But more importantly, probability of a hypothesis given the data is not the same as probability of the data given the hypothesis; we are usually more interested in the former. For example, the fact that probability of the data given a null hypothesis is low does not imply that probability of an alternative hypothesis is high, as it is often done. Imagine that we have got some extreme data. Our probability of the data given a null hypothesis can be low, but probability of the data given an alternative hypothesis can also be low, and it still says nothing about the probability of the hypothesis given the data. The latter probability, as we know from Bayes' theorem, depends on probability of the data *under all possible hypotheses*.

In the following sections I will focus on practical aspects of multivariate data analysis in community ecology, demonstrating the mentioned characteristics of both approaches in action.

[25] E.g., R package "pcaMethods" I used in section 6.3.2.2 uses EM algorithm, while "boral" package uses Gibbs sampler in JAGS. Even basic models of Bayesian PCA in "boral" take much longer time to compute than similarly complicated model with missing values evaluated in "pcaMethods". For some tasks, such as estimation of missing values, the Gibbs algorithm can be too slow.

6. Multivariate analysis

So far I have presented the differences between the classic and Bayesian approaches that arise from different interpretations of probability, or from different ways of doing inference. Although I used perspective of a community ecologist in the discussion of which of the differences have an impact on data analysis, until now the majority of the conclusions applied to all fields of science. From now on I will focus on typical data formats and techniques that are applied in community ecology, although not all of them are unique to community ecology. Apart from the description of the data and methods I will also present example analyses of real-world ecological data. I believe that they are crucial for putting the somewhat abstract arguments into practical context.

There are two basic approaches to obtain data: observational and manipulative (experimental). Either way, we gather (multivariate) data about some communities; these could be sown assemblages of species in pots, grasslands in a particular region, or succession plots on spoil tips across a country. In botany the typical tool is a phytosociological relevé (Mucina, Schaminée, & Rodwell, 2000) – a dataset of community composition at a small scale (e.g. 16 m^2), which serves as a representation of a larger community, and which I will use in the following sections.

6.1 Multivariate data

The multivariate community data can have several forms such as presence/absence, percentage cover[26], or number of individuals[27] of each species at each site (relevé) (Fig. 4). Plus, we can have data on environmental conditions at each site, for example data on temperature, soil

[26] Several sampling (mostly semi-quantitative) scales have been introduced and are widely used, with the typical example being the 7-grade Braun-Blanquet scale (Braun-Blanquet, 1964).
[27] Individuals in plant species are often hard to distinguish in field. For this reason, and because sometimes we are more interested in actual physically distinct units, we count ramets instead of genets.

nutrients, or on treatment identity in case of a manipulative study (Fig. 4). Usually the environmental data are used as predictors of the community data.

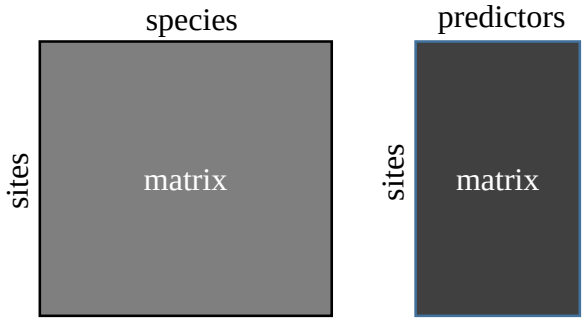

Fig. 4: Typical form of community ecology data, consisting of two matrices. On the left is a matrix of occurrence of species on sites and on the right is a matrix of predictor values on sites.

We usually ask two questions. The first is concerning only the left matrix – we want to know the structure of our communities – we are interested which species co-occur, which do not, and whether there are distinct gradients in community composition. This is assessed through covariance between species in the communities. This first question is also the central focus of the rest of this thesis. Second question, focusing on relationship between the two matrices, is: "Which predictors are associated with the species composition of the communities, and how strong is the association?"

When analyzing a dataset, the first step is usually a mere exploration – we plot histograms of variables, pondering the distributions they resemble, examining their range and other descriptive characteristics, and potentially exploring bivariate relationships. Such exploration helps to decide which methods of further analysis will be used, and is useful in later interpretation of the results (Tukey, 1977). However, such simple explorations miss potentially complex associations between more than two variables, so it would be practical to have an exploration method which would make the more complex structures of the multidimensional data visible.

6.2 Ordinations

Ordinations are group of techniques that enable to visualize complex multidimensional data in two dimensions. In Figure 5 I show graphs of species abundance data (from the left matrix from Fig. 4).

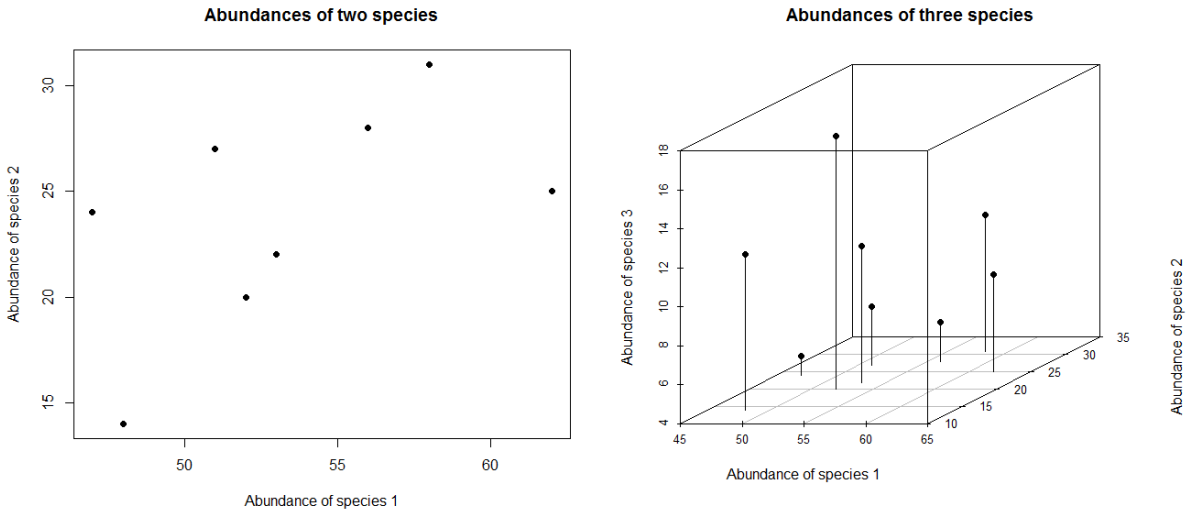

Fig. 5: Left: Abundances of two species. Plot of one species against the other is two dimensional. Right: Analogical plot for three species. Simulated data.

In case of three species (Fig. 5, right), we add another axis, so that the data are three-dimensional. The number of dimensions of the data is the same as the number of species at all sites, and the number of data-points dispersed in the multidimensional space correspond to the number of our sites. Our data form a multidimensional cloud of points and we are interested in its structure.

There are two ways to visualize a multidimensional dataset. First, we can project the data points in lower (typically two) dimensional space (Gauch, 1982). The projection can be done in several ways, but the central idea is the same: maintain as much information about the data as possible. This is done by preserving distances (e.g. Euclidean) among the data points. Examples are principal component analysis (PCA), canonical correspondent analysis (CCA), or principal coordinates analysis (PCoA) (Legendre & Legendre, 1998).

In an alternative approach we model the data, assuming that there are unmeasured variables which govern the variability of the data. Such latent variables can be used as axes of an ordination diagram.

6.2.1 Classic principal component analysis (PCA)

The first and perhaps simplest and most famous classic ordination technique is principal component analysis (PCA) (Jolliffe, 2002). It preserves Euclidean distances between the data points. PCA is a transformation method; if we imagine a cloud of data-points as static, PCA just rotates its coordinate system. The rotation is done so that the first axis goes through the most elongated part of the cloud. Second axis is orthogonal to the first one, and goes through

the second longest part of the cloud. Third axis is again orthogonal to the first two, and so on. In the end we have exactly the same multidimensional cloud of the data-points, only the coordinates of each point are different. Hence, we can see that the distances between the data-points have remained intact.

To visualize the re-projected dataset, we simply plot the data-points using the coordinates on the two axes of interest. For example, if we have data-point with coordinates (after transformation): [1,5,7,2,6,5] in a six-dimensional space, the projected data-point is going to have coordinates [1,5] on the first two axes. In practice, PCA is usually computed from eigenvalues of correlation matrix of the data (Legendre & Legendre, 1998).

In a meaningful PCA, all original variables have to be in the same units, otherwise they have to be standardized, since the most elongated direction of the data would be given purely (and arbitrarily) by the units. Then there is the problem of the distance measure. Species are thought (according to niche theory; (Hutchinson, 1957)) to have an optimum along environmental gradients. In optimum their abundances tend to be greatest and they decrease as we move away from the optimum in both directions. Hence, similar suboptimal abundance values can emerge at two different parts of the gradient. Hence, Euclidean distances based on abundance data along long gradients make no sense, since they would assign small projected distances between sites that are far apart on the gradient. This is known as a *problem of double-zeros*, and it is a common issue especially in abundance data (Legendre & Legendre, 1998), usually addressed by the use of correspondence analysis[28].

6.2.1.1 Example of classic PCA

For an example analysis I use data from Železné hory mountains (section 9.1). These are zero-inflated data on species abundances of plants at a set of sites, and on environmental conditions at the sites. Such zero-inflated abundance data are not suitable for principal component analysis (see previous section 6.2.1). Hence, I will do the PCA using the data on environmental conditions[29]. I standardized the data by columns (because columns are in different units; section 9.1).

[28] There is also possibility to discard rare species from our analysis. This way we weaken the problem of double-zeros. Discarding of rare species is permissible in case when we consider them as insufficiently sampled and their presence on particular localities/plots not informational. On the other hand, rare species could be regarded as important indicators of different environmental conditions among localities/plots.

[29] Or more precisely characteristics of localities which encompass soil nutrients, organic matter, pH, ground water level, biomass and area of localities. For simplicity, I will refer to them as to environmental conditions.

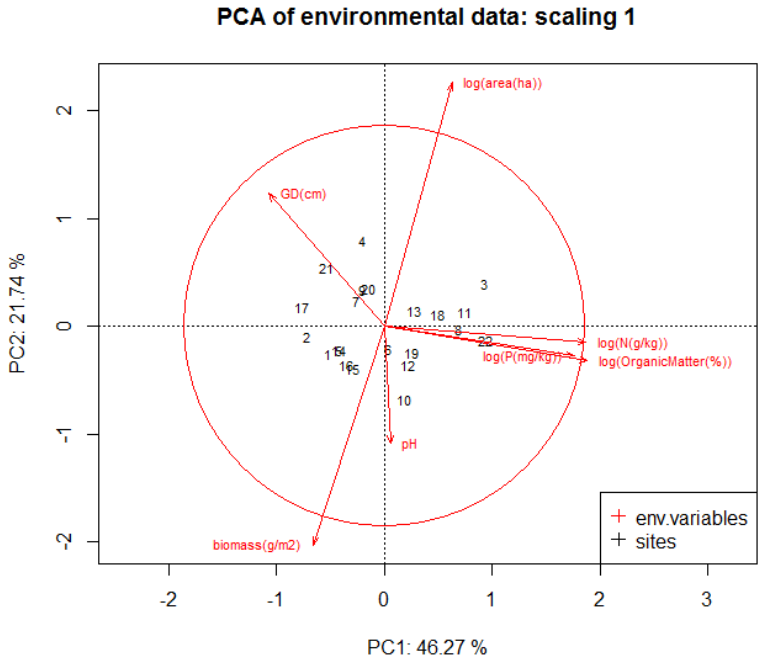

Fig. 6: Principal component analysis of environmental data using scaling 1. Sites are visualized as numbers. Descriptors as red arrows. Red circle is circle of equilibrium descriptor contribution. Descriptors reaching cross red circle significantly contribute to the formation of first two axes (Legendre & Legendre, 1998). *For data description see section 9.1.*

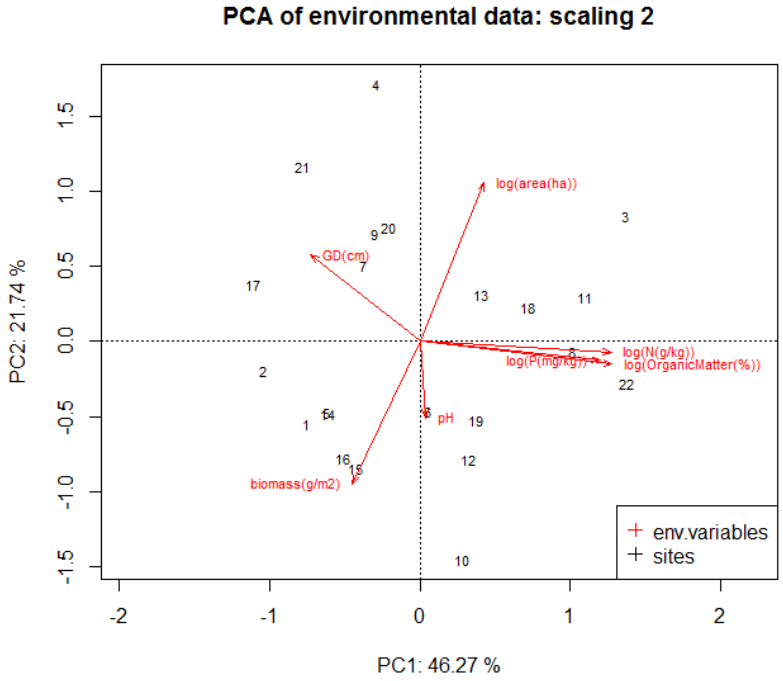

Fig. 7: Principal component analysis of environmental data using scaling 2. Sites are visualized as numbers. Descriptors as red arrows. For data description see section 9.1.

Figures 6 and 7 show the first two principal components. The two figures differ only in scaling of the results. The first scaling (Fig. 6) represents the orthogonal projection into two-

dimensional space. From that follows that the distances between the data-points in the figure are approximation of their Euclidean distances. If we project a data point on some of the descriptors, we get an approximate value of the environmental condition on the site – site 22 has one of the highest phosphorus contents and site 17 one of the lowest. Length of the arrows corresponds to their correlation with the first two principal axes. In contrast to Figure 6, Figure 7 shows the data in such a way that the angle between the predictors correspond to their covariance (Legendre & Legendre, 1998).

6.2.2 Bayesian PCA

Bayesian approach offers a possibility to execute an analysis that is similar to the classic PCA. It also enables to visualize the multidimensional data in a lower-dimensional space, and it gives similar results. However, the Bayesian PCA is not identical with the classic PCA (potentially raising doubts about the appropriateness of calling it PCA at all).

In the classic PCA we transform the points' coordinate system, and project them to lower dimensional space – it can be seen as a data transformation exercise. In contrast, the Bayesian PCA is model-based (see section 6.2), probabilistic (we assign probability to the data and the model parameters), and the data points are treated as outcomes of the model, whose parameters are then estimated from the data. The model is then visualized.

The Bayesian PCA is usually referred to as *latent variable model*[30] (hereafter LVM). It was introduced in 1997 (Tipping & Bishop, 1997), but was only recently provided for broad use in R-package[31] "boral" (Hui, 2015a), so it can be easily used by ecologists. The latent variable model approach models data using latent (unobserved) variables[32]. The data are assumed to be a linear transformation of these latent variables, with and added Gaussian error term (noise) (C M Bishop, 1999). The estimated latent variables are then plotted as axes of an ordination diagram, getting visual output which is similar to the classic PCA biplot.

6.2.2.1 Example of Bayesian PCA

For illustration of the Bayesian PCA, I used the same environmental condition data as in section 6.2.1.1 for the classic PCA. The analysis was done by R-package "boral" (Hui, 2015a). The

[30] We can do the latent variable model also in the classic approach. This would however miss some advantages of Bayesian approach, such as posterior distributions or automatic dimensionality determination (Christopher M Bishop, 1998). On the other hand, it will be less computationally demanding, making it potentially usable for more complicated tasks and larger datasets.

[31] All analyses in the thesis are done using program R (R Core Team, 2016).

[32] Bishop refers to one multivariate latent variable. We will use terminology of Warton who speaks about univariate latent variables. Meaning is identical.

model was defined as containing two latent variables, to be used for ordination in two dimensions. Distribution (error structure) of the data had to be defined, and I used normal distribution since the variables have approximately normal frequency distribution[33] (after logarithmic transformation in several cases).

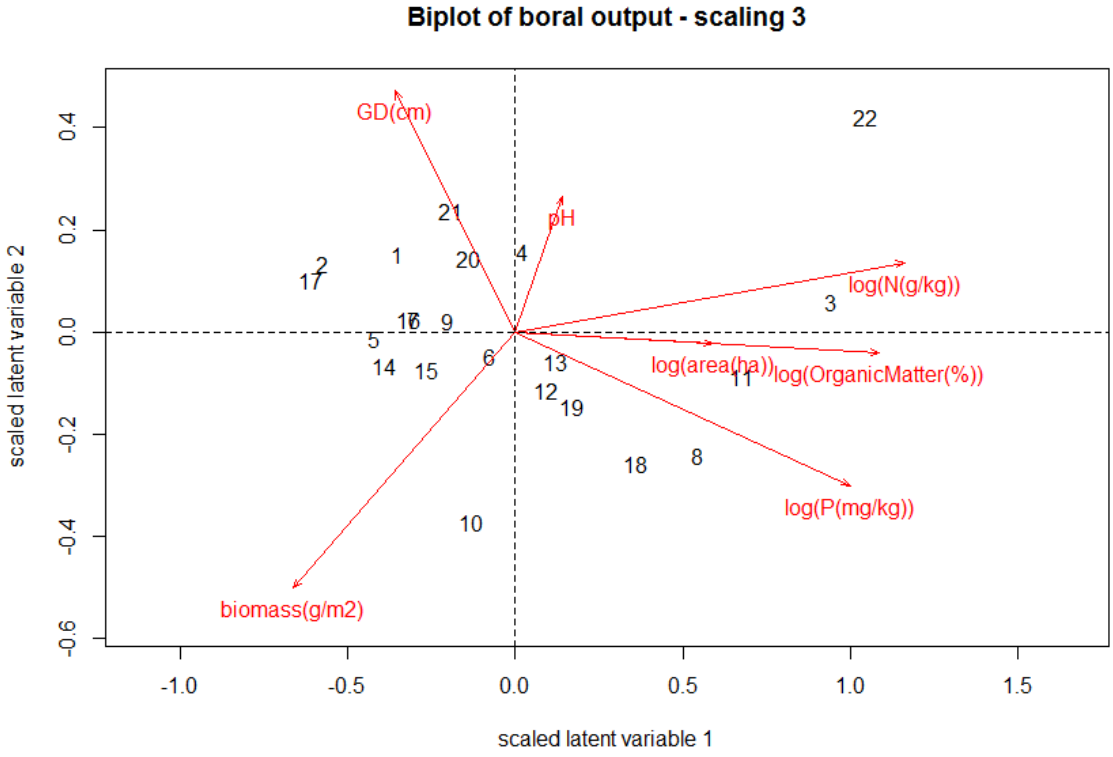

Fig. 8: Biplot of the latent variable model. Sites are visualized as numbers. Descriptors as red arrows. For data description see section 9.1.

Figure 8, the resulting Bayesian PCA biplot, is qualitatively similar to the one we have got from the classic PCA (Figure 6 and 7). We can see the nearly identical position of the environmental variables, and similar position of the data points. Scaling of both the environmental variables and the data points (scaling 3 (Oksanen et al., 2016)) is used in the visualization[34]. Note that a measure of explained variability by each axis is missing[35]; this is because a proper way of calculating percentage of explained variability is not settled yet (Hui

[33] Each used variable passed Shapiro-Wilk normality test, although some of them with p-value less than 0.1. We would get better results with a dataset which would perfectly match the assumed distributions (tested on in-build R dataset "Iris" where ordination plots from classic and Bayesian principal component analysis are hardly recognizable). I decided to use real data for illustration of the methods, in order to encounter patterns and problems that we could encounter in a real data analysis.

[34] On Figures 10 and 11 we can see how scaling 1 looks like.

[35] The scale of the second latent variable is narrower than the scale of the first one, meaning that the variability explained by the first axis is higher. Compare this with Fig. 10, showing the same results, where the scale of both axes is the same.

2016, personal communication). Importantly, interpretation of the results is the same as in classic PCA (Warton, Blanchet, et al., 2015). Latent variables represent both main axes of covariation between descriptors and missing (unmeasured) predictors.

6.2.2.2 Other tools for Bayesian PCA

So far we used R-package "boral"; this package offers more than a simple Bayesian PCA analog. Two things are worth mentioning: First, it can handle non-normal distributions, which are Binomial, Poisson, negative binomial, lognormal, tweedie, exponential, gamma, beta and ordinal (done as cumulative probit regression) (Hui, 2015b). Second, it enables to include predictor variables, so that we get analysis analogical to redundancy analysis (RDA; (Legendre & Legendre, 1998)) – part of the variability would be explained by our predictors and latent variables would then explain maximum possible amount of the residual variability.

Apart from "boral", there are several other tools for multivariate Bayesian data analysis[36]. Package "pcaMethods" (Stacklies, Redestig, Scholz, Walther, & Selbig, 2007) offers PCA methods including Bayesian PCA, and it can handle incomplete data (more details viz. section 6.3.2.2). Package "bPCA"[37] (Smycka & Keil, 2015) is alternative approach to Bayesian principal component analysis. It models data as outcomes of a multivariate normal model; parameters of the model (the vector of means and the covariance matrix) are estimated using MCMC, and classic PCA is then calculated using eigenvalues of the covariance matrix (which comes with quantified uncertainty about the covariance values). For broader list of other related tools see: (Warton, Foster, De´ath, Stoklosa, & Dunstan, 2015).

6.3 Comparison of the classic and Bayesian multivariate analysis

In the following section I will focus on differences between the classic and Bayesian approach in field of multivariate analysis. I will look for practical reasons for preferring Bayesian approach, apart from reasons related to its theoretical distinctions (discussed in section 5). First, I will present advantages of latent variable models in comparison with other model approaches. Then I will continue with advantages of Bayesian latent variable model.

[36] Mentioned are only R packages.
[37] Beware: There is also R package called „bpca", which however does not provide Bayesian principal component analysis.

6.3.1 Latent variable models (LVM) vs. multivariate generalized linear mixed models (GLMM)

So far I have dealt with the Bayesian PCA (see section 6.2.2) using the latent variable model (LVM; package "boral") and I provided comparison with the classic distance-based PCA (section 6.2). However, the LVM is not the only method suitable for Bayesian multivariate analysis. Latent variable model is an extension of generalized linear models (GLM), but there are other possible extensions.

The purpose of a multivariate model is to model abundances of multiple species on sites as response to predictor variables capturing correlation across these species[38](Warton, Foster, et al., 2015). Correlation can be introduced as a random effect applied to each site, but this leads to a positive and constant covariance between species, what is biologically unrealistic (Warton, Blanchet, et al., 2015). We can avoid this problem by introducing correlation via multivariate random effect at the site level (which is done in the so called GLMM). This is applicable in situations where we have large number of sites in comparison with number of our species, since number of parameters of such model increases quickly with the number of species in the data (Warton, Blanchet, et al., 2015). Often, this is not the case of community ecology data, the number of species is frequently not much lower than the number of sites, and often even bigger.

In contrast, a latent variable model (LVM) deals with among-species correlations using latent variables, and so it reduces the (potentially high) number of parameters which have to be estimated. As such it can be used for community data with more species than sites. Differences between these two approaches are illustrated in Figure 9.

[38] So called „joint model" (Warton, Blanchet, et al., 2015).

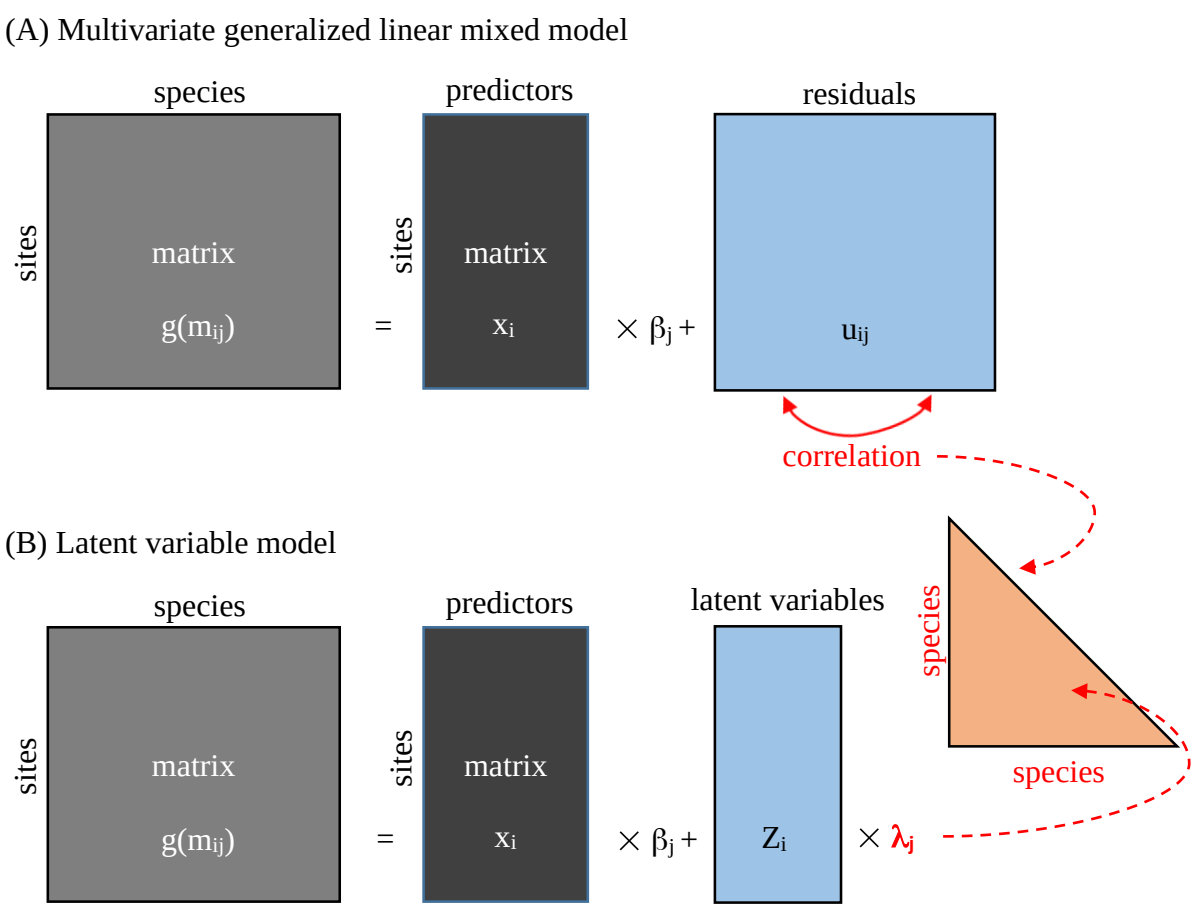

Fig. 9: : Schematic of GLMM and latent variable models (adapted from Warton et al. (2015)). Correlation between occurrences of species can be handled in different ways: (A) A multivariate GLMM uses correlated multivariate random effects, u_{ij}, to estimate correlation. (B) A LVM includes a smaller number of latent variables, z_i, which play the role of missing predictors. Their factor loadings, λ_j, approximate the correlation across species but use fewer parameters than the GLMM.

6.3.2 Advantages of Bayesian approach

Bayesian multivariate analysis has several advantages over the classic approach. These advantages follow both from the principles of Bayesian analysis (section 3), and from practical considerations, and are potentially key for predicting the impact of Bayesian thinking on the field of community ecology.

6.3.2.1 Advantage of explicit uncertainty

As I have mentioned (section 3), the output of Bayesian analysis provides not a single best estimate of a parameter, but a distribution of parameter values (more accurately: its probability density, or mass). From that distribution we can see the most probable value of the parameter (which we know from classic analysis) as well uncertainty around the parameter estimate. The multivariate Bayesian analysis is no exception – we can use the entire distribution of

parameters, and visualize it in ordination plots. Note: We can indeed use just a single expected value[39] of the parameters, which gives ordination biplots similar to the classic ordination methods.

The R's "boral" LVM provides standard deviation of posterior distributions, which can be used to construct an ordination plot with uncertainties – something that has not, to my knowledge, been shown in the literature. The output of the LVM contains the latent variables and their factor loadings; by their multiplication we get coordinates of sites, which are dispersed on a 2D plane (in case we used two latent variables). The plane is located in the same multidimensional space as were the sites before we have done LVM. For the purpose of plotting, the coordination system is rotated as in the classic PCA (as I said in section 6.2.1, PCA is just a rotation of coordination system, which is exactly what we need here).

To plot the uncertainties, I had to work with (product of) both the uncertainties of the latent variables and their loadings. Since the product of two normally distributed variables is not easily determined, numerical sampling (from the posterior distributions) can be used (Seijas-Macias & Oliveira, 2012), which is what I have done. Through sampling from the posterior distributions of latent variables and their factor loading, and then multiplication, I got points which were lying not only on the plane in n-dimensional space, but also around it. I then projected all sampled points on this plane (plane was the one mentioned earlier which comes from medians of posterior distributions of parameters of LVM). Finally, the plane was rotated via PCA for visualization.

Figures 10 and 11 show the resulting plots. The biplots are scaled in order to visualize the descriptors (Fig. 10) and sites (Fig. 11). They can be interpreted as classic ordination plots scaled as distance plots (scaling 1, see section 6.2.1.1). In addition, the uncertainties are displayed – red dots are the optimum values (medians of posterior distributions of parameters). Grey dots are sampled from the posterior distribution[40] of particular parameters and projected into the ordination plane. Clearly, the uncertainty is huge, grey dots are nearly everywhere.

Such result is striking, and casts doubt on whether we can actually derive anything useful and general from position of data points in an ordination plot based on data matrix of 22 rows and 7 columns. I suggest that this can be a serious problem of many ordination analyses that are done in a classic setting (e.g. Lepš and Šmilauer (2003)), and I argue that this issue should be further investigated, perhaps using a more extensive simulation study that would determine the sample size, number of species, and strength of gradients that are sufficient to

[39] Usually median of posterior probability distribution is used. Also mean is possible.
[40] Normal distribution with corresponding mean and standard deviation.

provide meaningful ordination diagrams. As a first hint, I added samples (black dots) from ten times narrower posterior distribution (Fig. 10 and 11), illustrating the magnitude of uncertainty reduction that we would need to be able to make confident inference from the biplot.

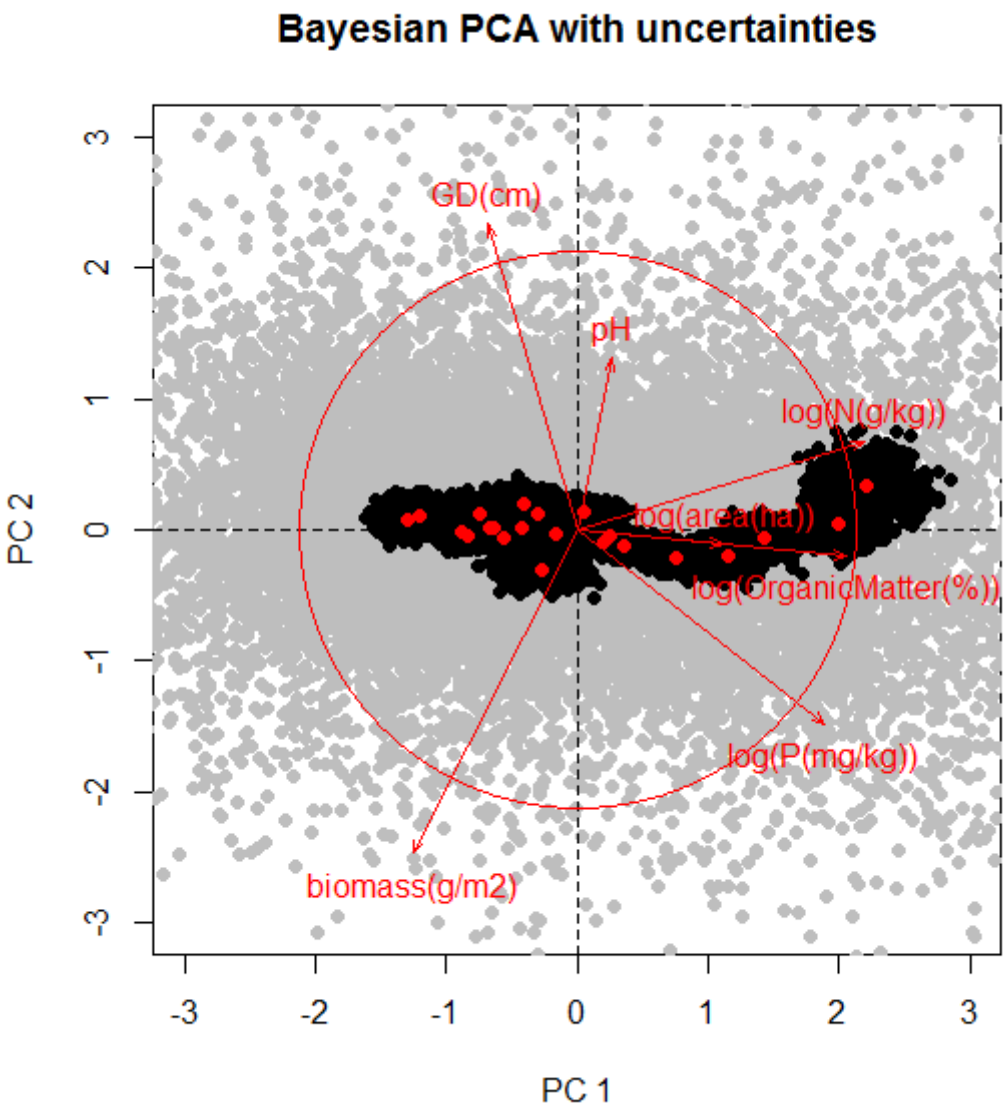

Fig. 10: Biplot from latent variable model with uncertainty. Sites (medians) are visualized as red points. Descriptors as red arrows. Descriptors reaching cross red circle are correlated with first two axes more than randomly. Grey points are 1000 of samples from distribution of particular parameters projected on the plane. They visualize uncertainty in position of red points on the ordination diagram. Black points are analogical to grey points; the difference is that they are sampled from a ten times narrower posterior distributions. For data description see section 9.1.

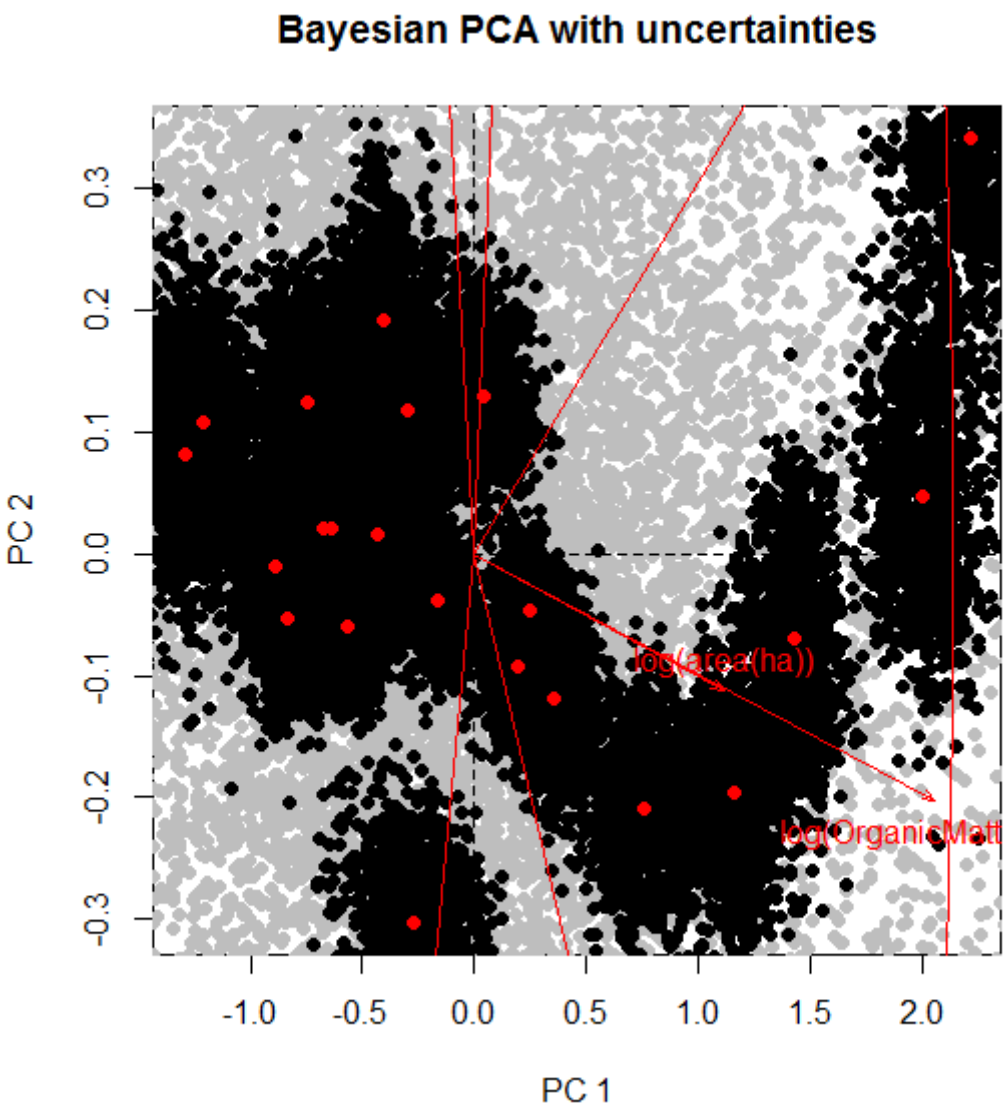

Fig. 11: Detail from Figure 10 focused on sites. Sites (medians) are visualized as red points. Descriptors as red arrows. Grey points are 1000 samples from posterior distribution of particular parameters, projected on the plane. They visualize uncertainty in position of red points on the ordination diagram. Black points are analogical to grey points, only sampled from ten times narrower distributions. For data description see section 9.1.

6.3.2.2 Advantage of meaningful treatment of missing values

Bayesian approach allows as to incorporate previous knowledge into our analysis, which is done via prior distributions of parameters which enter an analysis (see section 3). Thanks to this, Bayesian analysis works efficiently and logically with incomplete data (Gelman et al., 2004).

Incomplete (missing) data are a serious nuisance in community ecology (Leps & Smilauer, 2003)[41]. The classic methods in multivariate analysis usually employ ad-hoc solutions to the problem, none of them completely satisfactory (Leps & Smilauer, 2003): A common approach is to discard observations (cases, table rows) which have some missing values, causing loss of potentially valuable data. In cases of fragmented ecological data (which are not uncommon), we could lose almost the entire dataset. For other ad-hoc solutions see Lepš & Šmilauer (2003).

In the model-based approaches there are methods designed for statistical *imputation* of missing values (Gelman et al., 2004). In Bayesian analysis this imputation can be done iteratively – results from one estimation can be used through priors in another to improve the estimation (Oba et al., 2003).

To examine the effect of missing data on Bayesian PCA I used R package "pcaMethods". The multivariate dataset from Železné hory that I used so far is fortunately complete. Hence, I randomly discarded 1/7 of values and performed the Bayesian PCA. Results are shown in Figure 12; position of descriptors in the right picture is similar to the results we got from the classic principal component analysis (Fig. 6 and 7). Unfortunately, package "pcaMethods" does not report uncertainties of estimates; these would be greater in case of missing values.

[41] e.g.: Missing values can be an information which does not exist (e.g., part of an experiment has been destroyed) or information we did not manage to acquire for various possible reasons, for example due to technical issues or limited financial resources.

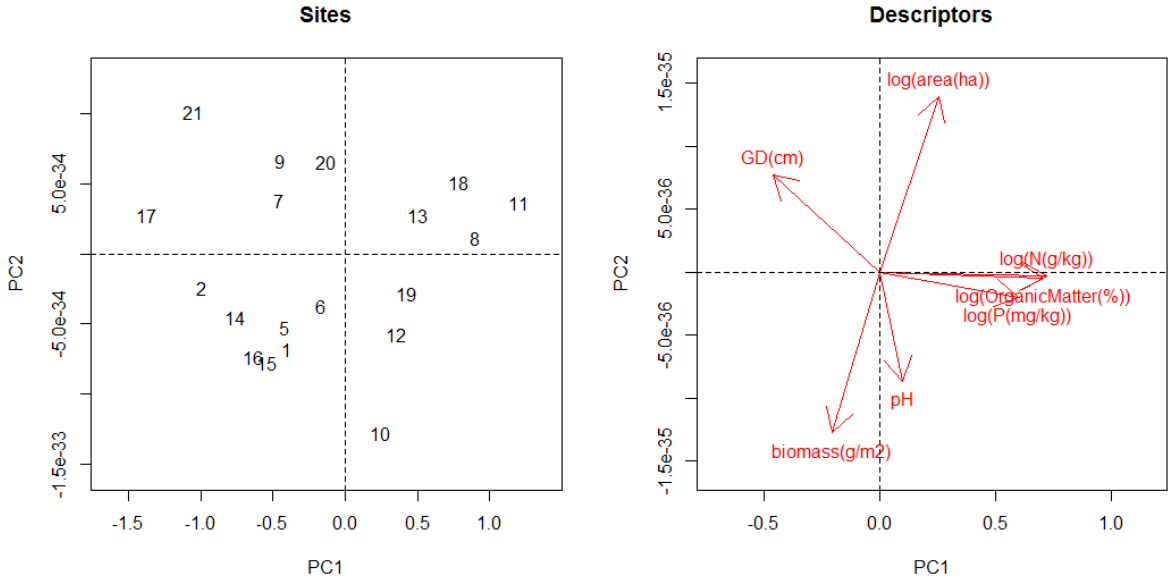

Fig. 12: Ordination plot from incomplete data. Sites are visualized by points on the left side. Descriptors are visualized by arrows on the right side. For data description see section 9.1.

6.3.2.3 Advantage of having a model

As I mentioned, the latent variable approach used in the Bayesian PCA falls in a group of model-based approaches to ordination analysis, which contrast to the most used methods which are essentially distance-based data transformations. The model-based approach can be used in the same way as other model-based techniques in statistical analysis: We can subject it to model selection (Burnham & Anderson, 2002), we can use it to make predictions (Warton, Blanchet, et al., 2015) and we can use it as building element in creation of more complicated models.

Model selection is usually done using some criteria which evaluate goodness of fit of the model, which is then penalized by the number of its parameters. Such criteria can be calculated in the case of Bayesian PCA. Package "boral" that we used for our analysis in section 6.2.2 provides several of them (for example: AIC, BIC).

Finally, when we have done a Bayesian PCA, we have a model which models our data. This model can be used to predict data that we have not measured. For example, I made a LVM model in "boral" of presence of our species on plots in Železné hory using environmental variables as predictors. We can now measure these environmental variables on other plots and use the model for prediction of their species composition and diversity.

I used cross-validation (Hastie, Tibshirani, & Friedman, 2009) to explore how good the prediction is. I used each time 21 localities to fit the predictive model and one locality to test the prediction. Results are not impressive in this case. The model successfully predicted species

diversity on 8 out of 22 localities (Fig. 13). For comparison, classic generalized linear model predicting species diversity directly (i.e., not modeling individual species) predicted 18 from 22 localities (Fig. 13). The reason for this result is probably the high number of rare species which occur only in several plots. This means that there is not enough information (signal) in the data that would enable to predict the rare species' occurrences based on the environmental gradients.

In spite of my unimpressive results, I suggest that predictive capabilities of the model-based ordination techniques should be further and systematically assessed, e.g. also through a simulation study. There has been a debate about whether it is better to model biodiversity directly as a single variable, or whether it is more advantageous to model individual species and stack their distributions (Guisan & Rahbek, 2011), and I see the model-based ordinations as a potentially new way to approach the problem.

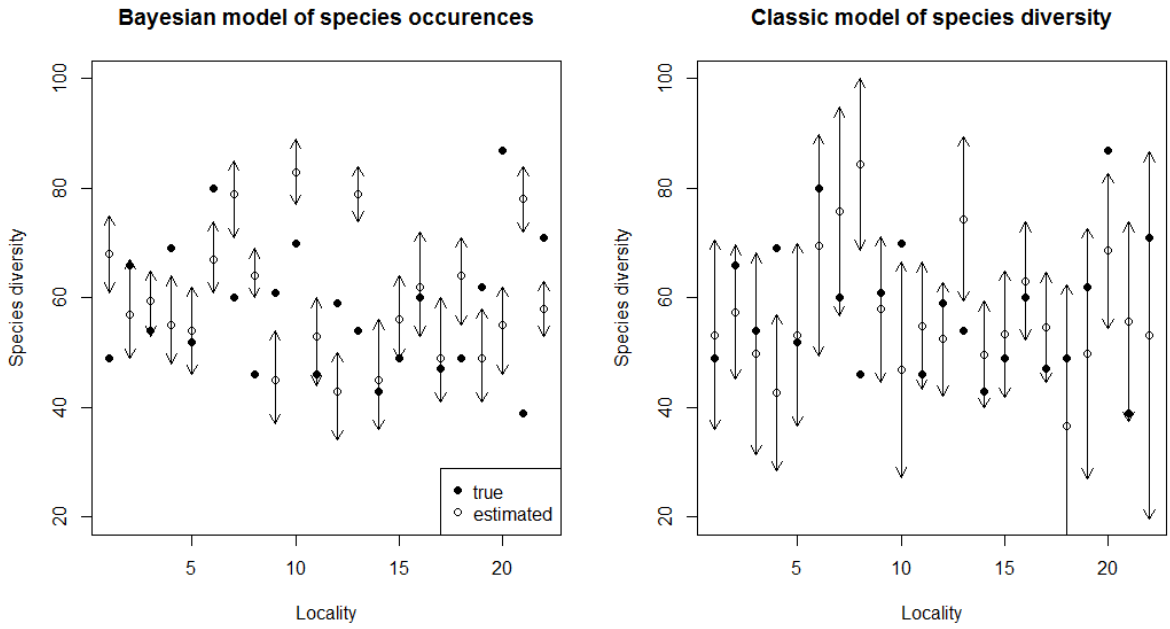

Fig. 13: Comparison of Bayesian model predicting occurrences of each species (left) and univariate classic model predicting species diversity (right). Bayesian model successfully predicted 8 localities, classic model predicted successfully 18 localities. Arrows indicate confidence or Bayesian intervals.

7. Conclusion

Bayesian approach brings new possibilities which are, thanks to the increasing power of computers, available to ecologists. There has been an especially rapid development in the field of multivariate analysis, and I hope that this thesis is a contribution to this development.

I have marked the way which lies in front of community ecologists who have currently been using classic statistical methods, and who might be tempted to get into methods of Bayesian analysis. These methods are presented not as an alternative, but rather as an enrichment of the classic tools. Ellison (2004) mentions that before using Bayesian analysis, one has to consider three epistemic differences between the classic and Bayesian approaches: the different interpretation of probability, the use of prior information, and the directly assigned probability of parameters. In section 5 I discussed all of them.

Bayesian approach has been challenged by several objections which target especially these issues, and which can prevent utilization of the approach. My aim has been to show to which degree these objections apply to data analysis in community ecology and how to address those which are relevant. First group of objections questions the subjective interpretation of probability. I argue that the answer to these objections is to adjust the subjective interpretation, so that it does not take our willingness to bet on outcomes literally.

Second group of objections aims at Bayesian inference: usage of priors and their specification has been criticized as subjective. My answer is that the specification of priors is just one of many subjective decisions scientists have to do in scientific process, and the efforts to eliminate them entirely is doomed to fail. Subjective decisions should be as transparent and open to discussion as possible, and specification of priors can indeed be done transparently. In problematic cases I suggested (section 3.1) that one possible solution is constraining of evaluated information – we construct priors to encompass only that information we know how to quantify as prior distributions. Furthermore, the process of specification of uninformative priors can be sometimes problematic (see Bertrand's paradox), but I do not see this as a concern for data analysis in community ecology, since in this field we rarely need to specify a single value of priors which would strongly depend on the model we have in mind. These answers could clear the way towards Bayesian approach. What remains is the call for broader adoption of the subjective interpretation of probability. I believe that this won't be a problem even for fans of the other interpretations, since different interpretations are potentially complementary and not necessarily exclusive.

Going through all of this and finally embracing Bayesian methods should have several advantages for community ecologists. In section 6.3.2 I illustrated the advantages related to ordination techniques, the discipline's core analytical toolset. The classic ordinations can be seen as mere data transformations. Bayesian latent variable models maintain the possibility of ordination for the purpose of data exploration, but in addition they also provide all advantages of parametric models – we can work with uncertainties, visualize them on ordination plots; we can do model selection and we can make predictions, and importantly, we can use them as building blocks of a more complex models. Multivariate generalized linear mixed models, which also have the advantages of parametric models, are another option. However, they are inappropriate for the typical community ecology data which have more species than sites, and I recommend the latent variable models as a more suitable alternative.

To conclude, I predict that Bayesian analysis will impact community ecology. The impact will likely to be associated more with practical aspects of Bayesian approach, rather than with theoretical underpinnings of probability. First, there will be the impact of the new methods. They pave an alternative way to deal with our data; and provide new perspective on how to solve problems. Second, with the ongoing discussions about what methods we should use and why (Jaynes, 1976), there will be greater focus on problematic parts of each approach, such as the stopping rule problem. This will help us to understand these problematic parts and hopefully eliminate mistakes which are done there. Third, Bayesian approach naturally integrates knowledge by incorporation of prior information, which could be especially important in making sense of the existing high volume published ordination analyses in community ecology – I suggest that such literature can provide valuable priors for future research. Finally, perhaps the largest impact will come from the explicit treatment of uncertainties, which could show that conclusions we have so far derived from classic ordination plots are often unjustified, since I have demonstrated that the (so far ignored) uncertainties are potentially huge.

8. Literature

Bayes, T., & Price, R. (1763). An Essay towards Solving a Problem in the Doctrine of Changes. *Philosophical Transactions*, (53), 370–418.

Bertrand, J. (1889). *Calcul des probabilités*. Gauthier-Villars.

Bishop, C. M. (1998). Bayesian PCA. *Advances in Neural Information Processing Systems (NIPS)*, *11*, 382–388.

Bishop, C. M. (1999). Variational principal components. In *9th International Conference on Artificial Neural Networks ICANN 99* (Vol. 1999, pp. 509–514). http://doi.org/10.1049/cp:19991160

Bolker, B. M. (2008). *Ecological Models and Data in R*. *Princeton University Press*. Retrieved from http://books.google.com/books?hl=en&lr=&id=7ZdtfhGHoz8C&oi=fnd&pg=PP2&dq=Ecological+Models+and+Data+in+R&ots=aZwx8lkksC&sig=r7LiufBUdrW_0t06xtTjy2JMeWU

Braun-Blanquet, J. (1964). *Pflanzensoziologie. Grundzüge der Vegetationskunde* (Third ed.). Wien: Springer.

Breslow, N. (1990). Biostatistics and Bayes. *Statistical Science*, *5*(3), 269–298. http://doi.org/10.1214/ss/1177012098

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. *Ecological Modelling* (2nd ed, Vol. 172). Springer. http://doi.org/10.1016/j.ecolmodel.2003.11.004

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, *12*(7), 693–715. http://doi.org/10.1111/j.1461-0248.2009.01314.x

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology Letters*, *8*(1), 2–14. http://doi.org/10.1111/j.1461-0248.2004.00702.x

Clark, J. S. (2007). *Models for Ecological Data*. *Princeton University Press* (Vol. 293). http://doi.org/10.1198/jasa.2008.s256

Crupi, V. (2015). Confirmation. In *The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2015/entries/confirmation/>*.

Dawid, A. P. (2004). Probability, Causality and the Empirical World: A Bayes–de Finetti–Popper– Borel Synthesis. *Statistical Science*, *19*(1), 44–57. http://doi.org/10.1214/088342304000000125

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, *7*(6), 509–520. http://doi.org/10.1111/j.1461-0248.2004.00603.x

Feyerabend, P. (1993). *Against Method* (Third Ed.). London, New York: Verso.

Gauch, H. G. (1982). *Multivariate Analysis in Community Ecology*. Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (Second Edi). Chapman & Hall/CRC.

Guisan, A., & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, *38*(8), 1433–1444. http://doi.org/10.1111/j.1365-2699.2011.02550.x

Hájek, A. (2012). Interpretations of Probability. In *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. *The*

Mathematical Intelligencer (2nd editio). Springer-Verlag. http://doi.org/10.1007/b94608
Havránek, T. (1993). *Statistika pro biologické a lékařské vědy*. Praha: Academia.
Horník, J., & Hrázský, Z. (2009). Polopřirozená travní společenstva jako významná refugia chráněných a ohrožených druhů rostlin – situace ve vybraný chmodelových úsecích okresu Chrudim [Semi-natural grasslands are important refuge for protected and endangered plant species: the situation . *Východočes. Sborn. Přírod. – Práce a Studie*, (16), 31–82.
Howson, C., & Urbach, P. (1991). Bayesian reasoning in science. *Nature*, *350*, 371–374.
Hui, F. K. C. (2015a). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in R. *Methods in Ecology and Evolution*, 1–7. http://doi.org/10.1111/2041-210X.12514
Hui, F. K. C. (2015b). Package "boral." R-package manual. Retrieved from https://cran.r-project.org/web/packages/boral/index.html
Hutchinson, G. E. (1957). Concluding remarks. In *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 22, pp. 415–427). http://doi.org/10.1101/SQB.1957.022.01.039
Jaynes, E. T. (1973). The Well-Posed Problem. *Foundations of Physics*, *3*(4), 477–492. http://doi.org/10.1007/BF00709116
Jaynes, E. T. (1976). Confidence intervals vs. Bayesian intervals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science* (Vol. II, pp. 175–257). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Confidence+intervals+vs+bayesian+intervals#0
Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition*. *Encyclopedia of Statistics in Behavioral Science* (Vol. 30). http://doi.org/10.2307/1270093
Joyce, J. (2008). Bayes' Theorem. In *he Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/>*.
Kemeny, J. G. (1955). Fair Bets and Inductive Probabilities. *The Journal of Symbolic Logic*, *20*(3), 263–273.
Klimešová, J., Janeček, S., Horník, J., & Doležal, J. (2011). Effect of the method of assessing and weighting abundance on the inter- pretation of the relationship between plant clonal traits and meadow. *Preslia*, *83*(i), 437–453.
Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*. (N. Morrison, Ed.) (Second Eng). Chelsea Publishing Company.
Laplace, P. S. M. de. (1902). *A Philosophical Essay on Probabilities*. *John Wiley & Sons* (Translated). http://doi.org/10.1017/CBO9781107415324.004
Lawton, J. H., & Kinne, O. (Ed. . (2000). *Excellence in Ecology: Community Ecology in a Changing World*.
Legendre, P., & Legendre, L. (1998). *Numerical ecology*. *Numerical Ecology Second English Edition* (Vol. 20). http://doi.org/10.1021/ic050220j
Leps, J., & Smilauer, P. (2003). *Multivariate Analysis of Ecological Data using CANOCO*. Cambridge University Press.
Lindley, D. V., & Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *The American Statistician*, *30*(3), 112–119. http://doi.org/10.2307/2683855
Lotka, A. J. (1910). Contribution to the Theory of Periodic Reactions. *The Journal of Physical Chemistry*, *14*(3), 271–274. http://doi.org/10.1021/j150111a004
McCarthy, M. A. (2007). *Bayesian Methods for Ecology* (Vol. 1). http://doi.org/10.1017/CBO9781107415324.004
McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community

ecology from functional traits. *Trends in Ecology and Evolution*, *21*(4), 178–185. http://doi.org/10.1016/j.tree.2006.02.002
Mucina, L., Schaminée, J. H. J., & Rodwell, J. S. (2000). Common data standards for recording relevés in field survey for vegetation classification. *Journal of Vegetation Science*, *11*(5), 769–772. http://doi.org/10.2307/3236581
Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *236*(767), 333–380. http://doi.org/10.1098/rsta.1937.0005
Oba, S., Sato, M. -a., Takemasa, I., Monden, M., Matsubara, K. -i., & Ishii, S. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, *19*(16), 2088–2096. http://doi.org/10.1093/bioinformatics/btg287
Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2016). vegan: Community Ecology Package. R package version 2.3-5. Retrieved from https://cran.r-project.org/package=vegan
Plato. (1883). *The Theaetetus of Plato*. (L. Campbell, Ed.)*Oxford University Press* (Second Edi, Vol. 1). http://doi.org/10.1017/CBO9781107415324.004
Popper, K. R. (1959). The Propensity Interpretation of Probability. *The British Journal for the Philosophy of Science*, *10*(37), 25–42.
Popper, K. R. (1983). *Realism and the Aim of Science*. Hutchinson.
R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/
Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
Reichenbach, H. (1971). *The Theory of Probability*. University of California Press.
Romeijn, J.-W. (2016). Philosophy of Statistics. In *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/statistics/>*.
Rosenkrantz, R. (1983). Why Glymour Is a Bayesian. In *Testing Scientific Theories, Earman, J. (ed.). Minnesota Studies in the Philosophy of Science (volume X., University of Minnesota Press, 1983)* (pp. 69–98).
Seijas-Macias, A., & Oliveira, A. (2012). An Approach to Distribution of the Product of Two Normal Variables. *Discussiones Mathematicae Probability and Statistics*, *32*(1–2), 87–99. http://doi.org/10.7151/dmps.1146
Simberloff, D. (2004). Community ecology: is it time to move on? *The American Naturalist*, *163*(6), 787–799. http://doi.org/10.1086/420777
Smycka, J., & Keil, P. (2015). Bayesian PCA. Retrieved July 31, 2016, from http://www.petrkeil.com/?p=2316
Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, (23), 1164–1167. Retrieved from https://github.com/hredestig/pcamethods
Tipping, M., & Bishop, C. (1997). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, *23*(3), 492–503. http://doi.org/10.1109/TNNLS.2012.2183006
Tukey, J. W. (1977). *Exploratory Data Analysis*. Pearson.
Van Dongen, S. (2006). Prior specification in Bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology*, *242*(1), 90–100. http://doi.org/10.1016/j.jtbi.2006.02.002
Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, *85*(2), 183–206. http://doi.org/10.1086/652373
Venn, J. (1876). *The Logic of Chance* (2nd editio). London: Macmillian and Co. Retrieved

from http://gdz.sub.uni-goettingen.de/dms/load/img/?PID=PPN309767121&physid=PHYS_0003
Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, *30*(12), 766–779. http://doi.org/10.1016/j.tree.2015.09.007
Warton, D. I., Foster, S. D., De′ath, G., Stoklosa, J., & Dunstan, P. K. (2015). Model-based thinking for community ecology. *Plant Ecology*, *216*(5), 669–682. http://doi.org/10.1007/s11258-014-0366-3

9. Appendix

9.1 Data description – Železné hory

The data come from an experiment conducted in Železné hory Mts in Czech Republic through years 2007-2013 (only data from year 2007 are used in this thesis). The data are frequencies[42] of individual species on 22 wet meadows spread across Železné hory Mts. The meadows were sampled in 2007 and then subjected to particular treatments and sampled again in the following years (Klimešová, Janeček, Horník, & Doležal, 2011).

9.1.1 Study area

The study area was located at an elevation from 340 to 550 m in Bohemia, Czech Republic (Klimešová et al., 2011). The landscape consists of a mosaic of temperate forests, arable land, intensively managed meadows and urban areas. Semi-natural, species-rich wet meadows occupy only fraction of the area, but host majority of the endangered plant species (Horník & Hrázský, 2009). For the experiment 22 wet meadows were selected, which are distributed over an area of 9 × 21.5 km, covering a range of environmental gradients (soil moisture, fertility, soil reaction values and types of management).

9.1.2 Experiment design

On each of the 22 meadows there were 4 permanent plots of 2 x 2 m separated by a 0.5 m buffering zone, resulting in 4.5 x 4.5 m blocks. Plant frequency was assessed in central 1 x 1 m quadrate of plots (recorded in the first half of July 2007). Full factorial design of treatment – mowing and fertilization – was applied. Mowing was done after vegetation assessment and then following year in the same time. Fertilization consists from application of 20g of fertilizer (mineral NPK: 10% N, 10% P_2O_5,10% K_2O) per m^2 at the end of July 2007 and 50g per m^2 in the second half of April 2008 and 2009. In the first half of July 2009 data on the response to the short period management were collected.

9.1.3 Dataset

The dataset consists of frequencies of each plant species in the 22 meadows in each of the 4 plots. Several characteristics of each meadow were also measured, these were: biomass [g/cm^2] for both years in plots which were mowed, underground water depth [cm] and content of some

[42] Assessed as occurrences (presence/absence) of species in 100 subplots of each plot.

nutrients in soil (NH_4, NO_3, PO_4). These data (characteristics) are available in Klimešová et al. (2011).

9.2 R-code

9.2.1 Stopping rule problem example

```
#stopping rule problem example
#first stopping rule
binom.test(40,100,p=0.5)

r_samples<-rbinom(10000,100,0.5)
hist(r_samples, freq=F, main="Histogram of number of successes in 100
trials with 0.5 probability success", xlab="Number of successes")
abline(v=40, col="red", lwd=2)
rect(30,0,40,0.08, col=rgb(0,0,1, alpha=0.2), border=0)
rect(60,0,70,0.08, col=rgb(0,0,1, alpha=0.2), border=0)
x <- 30:70
lines(x, dbinom(x, 100, 0.5),col="purple", lwd=2)

sum(r_samples<41)/10000*2     #p-value
sum(r_samples==40)/10000      #probability of exactly 40 successes

#second stopping rule
a_sample<-vector("numeric",10000)
for (j in 1:10000){
      a_sum <- 0
      i <- 0
      repeat {
            a_try<-rbinom(1,1,0.5)
            a_sum <- a_sum+(0.5>a_try)
            i <- i+1
            if (a_sum >= 60) break
      }
      a_sample[j]<-i
}
a_success<-a_sample-60
hist(a_success, freq=F, main="Histogram of number of successes in trials
with 0.5 probability success", xlab="Number of successes")
abline(v=40, col="red", lwd=2)
rect(10,0,40,0.04, col=rgb(0,0,1, alpha=0.2), border=0)
rect(80,0,150,0.04, col=rgb(0,0,1, alpha=0.2), border=0)

sum(a_success<=40)/10000*2    #p-value
```

9.2.2 PCA: classic and Bayesian

```
#PCA: classic and Bayesian
library(vegan)
```

```
library(boral)

#data preparation and functions

summary(data_pca)       #environmental data (Klimesova et al. 2011)
data_pca_st<-decostand(data_pca,"standardize")

"pcacircle" <- function (pca)
{
# Draws a circle of equilibrium contribution on a PCA plot
# generated from a vegan analysis.
# vegan uses special constants for its outputs, hence
# the 'const' value below.
# Authors of this function: Francois Gillet & Daniel Borcard, 24 August
2012

      eigenv <- pca$CA$eig
      p <- length(eigenv)
      n <- nrow(pca$CA$u)
      tot <- sum(eigenv)
      const <- ((n - 1) * tot)^0.25
      radius <- (2/p)^0.5
      radius <- radius * const
      symbols(0, 0, circles=radius, inches=FALSE, add=TRUE, fg=2)
}

#classic PCA
pca_out<-rda(data_pca_st)

#scaling 1
pc1<-round(pca_out$CA$eig[1]/sum(pca_out$CA$eig),4)*100
pc2<-round(pca_out$CA$eig[2]/sum(pca_out$CA$eig),4)*100
plot(pca_out, scaling=1, type="n", main="PCA of environmental data: scaling
1", xlab=paste("PC1:",pc1,"%"), ylab=paste("PC2:",pc2,"%"))
text(pca_out, scaling=1, display="si", pch=16, cex=0.7,
labels=as.character(c(1:22)))
text(pca_out, scaling=1, display="sp", cex=0.7, pos=c(4,2,4,2,4,4,4),
col="red")
pcacircle(pca_out)
legend("bottomright", pch=3, col=c("red","black"),
legend=c("env.variables","sites"))
p <- length(pca_out$CA$eig)
spe.sc1 <- scores(pca_out, display="sp", scaling=1, choices=c(1:p))
arrows(0, 0, spe.sc1[,1], spe.sc1[,2], length=0.07, angle=20, col="red")

#scaling 2
plot(pca_out, scaling=2, type="n", main="PCA of environmental data: scaling
2", xlab=paste("PC1:",pc1,"%"), ylab=paste("PC2:",pc2,"%"))
text(pca_out, scaling=2, display="si", pch=16, cex=0.7,
```

```
labels=as.character(c(1:22)))
text(pca_out, scaling=2, display="sp", cex=0.7, pos=c(4,2,4,2,4,4,4),
col="red")
legend("bottomright", pch=3, col=c("red","black"),
legend=c("env.variables","sites"))
p <- length(pca_out$CA$eig)
spe.sc1 <- scores(pca_out, display="sp", scaling=2, choices=c(1:p))
arrows(0, 0, spe.sc1[,1], spe.sc1[,2], length=0.07, angle=20, col="red")

#Bayesian PCA
bpca_out<-boral(data_pca_st,family="normal",num.lv=2)
plot(bpca_out)

bpca.plot <- function(bpca_out, scal)
{
# bpca_out:       model; output of function boral
# scal:           scaling of visualized data
      tcov <- bpca_out$lv.median %*% t(bpca_out$lv.coefs.median[, 2:3])
      pca_tcov <- rda(tcov)

      plot(scores(pca_tcov, scaling=scal)$sites[,1],scores(pca_tcov,
      scaling=scal)$sites[,2], type="n", xlim=range(c(scores(pca_tcov,
      scaling=scal)$sites[,1]-0.5,scores(pca_tcov,
      scaling=scal)$species[,1]+0.5)), ylim=range(c(scores(pca_tcov,
      scaling=scal)$sites[,2]-0.2,scores(pca_tcov,
      scaling=scal)$species[,2])), main=paste("Biplot of boral output -
      scaling ",scal), xlab="scaled latent variable 1", ylab="scaled latent
      variable 2")
      text(scores(pca_tcov, scaling=scal)$sites[,1],scores(pca_tcov,
      scaling=scal)$sites[,2])
      arrows(0,0,scores(pca_tcov,
      scaling=scal)$species[,1],scores(pca_tcov, scaling=scal)$species[,2],
      col="red", length=0.07, angle=20)
      text(scores(pca_tcov, scaling=scal)$species[,1],scores(pca_tcov,
      scaling=scal)$species[,2], col="red", labels=names(pca_tcov$colsum),
      pos=1)
      abline(h=0, lty=2)
      abline(v=0, lty=2)
}

bpca.plot(bpca_out, 3)
```

9.2.3 Bayesian PCA with uncertainties

```
#bayesian component analysis plot with uncertainties
library(vegan)
library(boral)
library(plotrix)

#data preparation
```

```
summary(data_pca)       #environmental data (Klimesova et al. 2011)
data_pca_st<-decostand(data_pca,"standardize")

#bpca
bpca_out<-boral(data_pca_st,family="normal",num.lv=2)

bpca.plot.unc <- function(bpca_out, sd_coe=1, n=1000,
col.points=rgb(0,0,0,alpha=0.2))
{
# bpca_out:       model; output of boral function
# sd_coe:         coeficient which multiplies uncertainties (1=real
uncertainties)
# n:              number of samples used to illustrate uncertainties
# col.points:     color of points representing uncertainties

tcov <- bpca_out$lv.median %*% t(bpca_out$lv.coefs.median[, 2:3])
pca_tcov <- rda(tcov)

sim<-array(dim=c(bpca_out$n,2,n))
sim_coe<-array(dim=c(bpca_out$p,2,n))
for (j in 1:2){
      for (i in 1:bpca_out$n){
            sim[i,j,1:n]<-rnorm(n, bpca_out$lv.median[i,j],
            bpca_out$lv.sd[i,j]*sd_coe)
}}
for (j in 2:3){
      for (i in 1:bpca_out$p){
            sim_coe[i,j-1,1:n]<-rnorm(n, bpca_out$lv.coefs.median[i,j],
            bpca_out$lv.coefs.sd[i,j]*sd_coe)
}}
tcov_sim<-array(dim=c(bpca_out$n,bpca_out$p,n))
for (i in 1:n){
      tcov_sim[,,i] <- sim[,,i]%*%t(sim_coe[,,i])
}
tcov_sim_sub<-array(dim=c(bpca_out$n,bpca_out$p,n))
for (i in 1:bpca_out$n){
      for (j in 1:bpca_out$p){
            for (k in 1:n){
                  tcov_sim_sub[i,j,k]<-tcov_sim[i,j,k]-tcov[i,j]
}}}
gua<-array(dim=c(bpca_out$n,bpca_out$p,2))
for (i in 1:bpca_out$n){
      for (j in 1:bpca_out$p){
            gua[i,j,1:2]<-quantile(tcov_sim_sub[i,j,1:n],c(0.025,0.975))
}}
colnames(gua)<-colnames(tcov)
gua_pca<-rda(gua[,,1])
gua_pca2<-rda(gua[,,2])
tcov_sim_pru<-array(dim=c(bpca_out$n,bpca_out$p,n))
```

```
for (i in 1:bpca_out$p){
      for (k in 1:n){
            tcov_sim_pru[,i,k]<-(tcov_sim[,i,k]-mean(tcov[,i]))
}}

tcov_c<-array(dim=c(bpca_out$n,bpca_out$p))
for (i in 1:bpca_out$p) tcov_c[,i]<-tcov[,i]-mean(tcov[,i])

plot((tcov_c%*%(pca_tcov$CA$v))[,1],(tcov_c%*%(pca_tcov$CA$v))[,2], asp=1,
col="red", pch=16, xlim=c(-3,3),ylim=c(-3,3), main="Bayesian PCA with
uncertainties", xlab="PC 1", ylab="PC 2")
abline(h=0, lty=2)
abline(v=0, lty=2)
for (i in 1:n) {
      points((tcov_sim_pru[,,i])%*%(pca_tcov$CA$v)[,1],(tcov_sim_pru[,,i])%
      *%(pca_tcov$CA$v)[,2], col=col.points, pch=16)
}
points((tcov_c%*%(pca_tcov$CA$v))[,1],(tcov_c%*%(pca_tcov$CA$v))[,2],
col="red", pch=16)
cons<-4
arrows(0,0,cons*pca_tcov$CA$v[,1],cons*pca_tcov$CA$v[,2], col="red",
length=0.07, angle=20)
text(cons*pca_tcov$CA$v[,1],cons*pca_tcov$CA$v[,2], col="red",
labels=names(pca_tcov$colsum), pos=c(3,1,3,1,1,3,3))
draw.circle(0,0,((2/bpca_out$p)^0.5)*cons, border="red")
}

bpca.plot.unc(bpca_out)
```

9.2.4 PCA with missing values

```
#PCA with missing values
library(vegan)
library(pcaMethods)

#data preparation
summary(data_pca)       #environmental data (Klimesova et al. 2011)
data_pca_st<-decostand(data_pca,"standardize")

#PCA without missing values
pcaM_out<-pca(data_pca_st, method="bpca")
slplot(pcaM_out)

#PCA with missing values
data_pca_st_m<-data_pca_st
while (sum(is.na(data_pca_st_m))<22){
      mm<-cbind(trunc(runif(1, 1,23)),trunc(runif(1, 1,8)))
      is.na(data_pca_st_m[mm[1,1],mm[1,2]])<-T}

pcaM_out_m<-pca(data_pca_st_m, method="bpca")
```

```
slplot(pcaM_out_m)
```

9.2.5 Cross-validation

```
#cross-validation
library(vegan)
library(boral)
library(mvtnorm)

#data preparation
summary(data_pca)       #environmental data (Klimesova et al. 2011)
summary(mat_pt3)        # presence/absence data of particular species in
particular locality

#Function predicting species diversity
predict_spp_rich<- function(mod_sampled, mat_pred, new_predictor=NULL){
# mod_sampled:    model, output of function "boral"
# mat_pred:       matrix of occurences of species on localities, which we
are trying to predict
# new_pred:       information about predicted localities, used as
predictors
# as basis for this function is used function published by Warton et al.
(2015).

true.spp.rich <- apply(mat_pred>0,1,sum)
sim_y <- matrix(NA,nrow(mat_pred),ncol(mat_pred))
sim_druh_bohatost <- matrix(NA,nrow(mat_pred),1000)
for (t in 1:1000){
      lv_sampled <- rmvnorm(nrow(mat_pred),rep(0,mod_sampled$num.lv))
      mean_resp <- as.matrix(cbind(1,lv_sampled,
      new_predictor))%*%as.matrix(t(cbind(mod_sampled$lv.coefs.median[,1:(1
      +mod_sampled$num.lv)],mod_sampled$X.coefs.median)))
      for (j in 1:ncol(mean_resp)) {
            sim_y[,j] <- rbinom(nrow(mat_pred), 1, p=pnorm(mean_resp[,j]))}
      sim_druh_bohatost[,t] <- rowSums(sim_y>0)
}
sim_druh_bohat_med <- apply(sim_druh_bohatost,1,median)
dol_qua<-vector("numeric",nrow(sim_druh_bohatost))
hor_qua<-vector("numeric",nrow(sim_druh_bohatost))
for (i in 1:nrow(sim_druh_bohatost)) {
      dol_qua[i] <- quantile(sim_druh_bohatost[i,],probs=0.025)
      hor_qua[i] <- quantile(sim_druh_bohatost[i,],probs=0.975)
}

out_sim <- rbind(true.spp.rich,sim_druh_bohat_med, dol_qua, hor_qua)
rownames(out_sim) <- c("true","estimated", "0.025", "0.975")
return(out_sim)}
#_____________________________
#cross-validation
vysledky<-matrix(NA,nrow=4,ncol=22)
```

```
for (test_id in 1:22){
      mat_sampled<-mat_pt3[-test_id,]
      mat_pred<-mat_pt3[test_id,]
      data_pca_sam<-data_pca[-test_id,]
      mod_sampled2<-boral(mat_sampled, data_pca_sam, num.lv=2,
      family="binomial")
      new_pred_1<- data_pca[test_id,]
      rownames(vysledky)<-c("true","estimated", "0.025", "0.975")
      vysledky[1:4,test_id]<-predict_spp_rich(mod_sampled2,
      t(matrix(mat_pred)), new_pred_1)
      print(test_id)
}
vysledky

#prediction with univariate glm model – for comparison
druh_bohat <- apply(mat_pt3,1,sum)
mod_1 <-
glm(druh_bohat~data_pca[,1]+data_pca[,2]+data_pca[,3]+data_pca[,4]+data_pca
[,5]+data_pca[,6]+data_pca[,7])
plot(mod_1)

pred_values <- vector("numeric",22)
pred_se <- vector("numeric",22)
for (i in 1:22){
      p1 <- data_pca[-i,1]
      p2 <- data_pca[-i,2]
      p3 <- data_pca[-i,3]
      p4 <- data_pca[-i,4]
      p5 <- data_pca[-i,5]
      p6 <- data_pca[-i,6]
      p7 <- data_pca[-i,7]
      mod_2 <- glm(druh_bohat[-i]~p1+p2+p3+p4+p5+p6+p7)
      pred_values[i] <- predict(mod_2, newdata=data.frame(p1=data_pca[i,1],
      p2=data_pca[i,2], p3=data_pca[i,3], p4=data_pca[i,4],
      p5=data_pca[i,5], p6=data_pca[i,6], p7=data_pca[i,7]),
      type="response")
      pred_se[i] <- predict(mod_2, newdata=data.frame(p1=data_pca[i,1],
      p2=data_pca[i,2], p3=data_pca[i,3], p4=data_pca[i,4],
      p5=data_pca[i,5], p6=data_pca[i,6], p7=data_pca[i,7]),
      type="response", se.fit=T)$se.fit
}
data.frame(druh_bohat,pred_values, pred_values+1.96*pred_se,pred_values-
1.96*pred_se)
```